

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО**

Факультет інформатики та обчислювальної техніки

(назва факультету, інституту)

Кафедра автоматизованих систем обробки інформації і управління

(назва кафедри)

"На правах рукопису"

УДК 004.021

«До захисту допущено»

Завідувач кафедри

О.А.Павлов

(підпис)

(ініціали, прізвище)

“ ” 20 18 р.

МАГІСТЕРСЬКА ДИСЕРТАЦІЯ

на здобуття ступеня магістра

за спеціальністю 126 Інформаційні системи та технології

(код та назва спеціальності)

ОПП

Інформаційні управляючі системи та технології

(код та назва спеціалізації)

на тему: Інформаційна технологія кластеризації користувачів

для підготовки цільових пропозицій

Виконав: студент

VI курсу групи ІС-72мп

(шифр групи)

Касянчик Дмитро Олександрович

(прізвище, ім'я, по батькові)

(підпис)

Науковий керівник

ст. викл. Олійник Ю.О.

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

(підпис)

Консультант

к.т.н., доц. Жданова О.Г

(науковий ступінь, вчене звання, прізвище, ініціали)

(підпис)

Рецензент

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

(підпис)

Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів без
відповідних посилань.

Студент

(підпис)

Київ – 2018

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ім. ІГОРЯ СІКОРСЬКОГО»**

Факультет інформатики та обчислювальної техніки
(повна назва)

Кафедра автоматизованих систем обробки інформації та управління
(повна назва)

Рівень вищої освіти другий (магістерський) за освітньо-професійною програмою

Спеціальність 126 Інформаційні системи та технології
(код і назва)

ОПП Інформаційні управляючі системи та технології
(код і назва)

ЗАТВЕРДЖУЮ
Завідувач кафедри

(підпис) О.А.Павлов
(ініціали, прізвище)
«__» грудня 2018 р.

ЗАВДАННЯ

на магістерську дисертацію студенту
Касянчику Дмитру Олександровичу
(прізвище, ім'я, по батькові)

1. Тема дисертації Інформаційна технологія кластеризації користувачів для підготовки цільових пропозицій
- науковий керівник дисертації ст. викл. Олійник Юрій Олександрович
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
- затверджені наказом по університету від “ 07 ” 11 20 18 р. № 4112с
2. Строк подання студентом дисертації “ 03 ” 12 20 18 р.
3. Об'єкт дослідження процес об'єднання користувачів в кластери за спільними інтересами, та використання цих кластерів для надання цільових пропозицій
4. Перелік завдань, які потрібно розробити проаналізувати існуючі засоби отримання даних про користувачів з соціальних мереж: facebook, twitter;
зібрати дані про вподобання користувача та його діяльність в соціальній мережі facebook;
обрати найбільш підходящий метод підготовки цільових пропозицій;
обрати та модифікувати метод кластеризації, для його використання в системі;
реалізувати алгоритм підготовки цільових пропозицій на основі обраних методів.

5. Орієнтовний перелік ілюстративного матеріалу ER діаграма; математична модель методу підготовки цільових пропозицій користувачам з використанням нечіткої кластеризації; схема алгоритму підготовки цільових пропозицій користувачам з використанням нечіткої кластеризації; схема алгоритму роботи модифікованого методу кластеризації на основі методу поширення схожості; результати експерименту; математична модель методу підготовки цільових пропозицій користувачам з використанням нечіткої кластеризації.

6. Орієнтовний перелік публікацій дві публікації тез доповідей на наукових конференціях «ІСТУ-2018» та «МОДС-2018».

7. Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

8. Дата видачі завдання “ 29 ” 10 20 18 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Строк виконання етапів магістерської дисертації	Примітка
1	Систематизація результатів огляду літератури	02.11.2018	
2	Порівняльний аналіз існуючих методів розв'язання задачі	05.11.2018	
3	Постановка та формалізація математичної моделі задачі	10.11.2018	
4	Модифікація існуючих методів розв'язання задачі	15.11.2018	
5	Розробка інформаційного та програмного забезпечення	22.11.2018	
7	Проведення експериментальних досліджень розроблених алгоритмів	26.11.2018	
8	Оформлення документації	29.11.2018	
9	Подання роботи на попередній захист	30.11.2018	
10	Подання роботи на основний захист	03.12.2018	

Студент

(підпис)

Д.О. Касянчик

(ініціали, прізвище)

Науковий керівник

(підпис)

Ю.О. Олійник

(ініціали, прізвище)

РЕФЕРАТ

Магістерська дисертація: 94 с., 21 рис., 32 табл., 1 додаток, 30 джерел.

Актуальність. Кожен із нас щодня отримує досить велику кількість інформації. Не завжди отримана інформація є релевантною, що зумовлює необхідність пошуку джерел, які будуть цікавими та яким можна довіряти. Більшість людей мають роботу, також багато хто має якісь захоплення. В обох випадках перед особою може постати задача у вивченні чогось нового, в поглибленні знань в певній сфері. Також майже кожен щодня дивиться фільми або читає книги, або ж прослуховує музику. І відповідно щоб виконати ці дії перед тим потрібно знайти матеріал, який буде цікавим. Досить часто особа тоне в тій величезній кількості матеріалів, коли шукає, саме те що їй потрібно. Адже до прикладу ми можемо знайти до 10, а інколи і більше книг на одну тему, а перечитати всі в нас просто не вистачить часу.

В описаних випадках виникає необхідність спитати в когось поради, і звісно не завжди в нас є така можливість. В таких випадках на допомогу, приходять саме рекомендаційні системи.

Більшість рекомендаційних систем видають пропозиції на основі раніше переглянутих або на основі схожих матеріалів. Ще одною досить вагомою проблемою таких систем є спрямованість лише на якусь одну тему. Дані системи часто не справляються з роллю порадника, коли особу цікавлять декілька тем. Також у випадку, якщо людина хотіла б дізнатися, що цікаво її однодумцям, то рекомендація на основі раніше переглянутих матеріалів може їй просто не підійти

Тому доцільною є розробка технології для об'єднання однодумців на основі діяльності в соціальних мережах. Завдяки застосуванню технологій з науки про дані система буде мати можливість кластеризувати користувачів зі схожими інтересами, що надасть можливість рекомендувати користувачам те, що входить в його коло інтересів.

Зв'язок роботи з науковими програмами, планами, темами. Робота виконувалась на кафедрі автоматизованих систем обробки інформації та управління

Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського» в рамках теми «Методи та технології високопродуктивних обчислень та обробки надвеликих масивів даних». Державний реєстраційний номер 0117U000924.

Мета дослідження – поліпшення якості та швидкості підготовки цільових пропозицій за рахунок удосконалення методу кластеризації користувачів..

Для досягнення мети необхідно виконати наступні завдання:

- проаналізувати існуючі засоби отримання даних про діяльність користувачів з наступних соцмереж: facebook, twitter;
- зібрати дані про вподобання користувача та його діяльність в соціальній мережі facebook;
- обрати найбільш підходящий метод підготовки цільових пропозицій;
- обрати та модифікувати метод кластеризації, для найкращого його використання в системі, що розробляється;
- реалізувати алгоритм підготовки цільових пропозицій на основі обраних методів.

Об’єкт дослідження – процес об’єднання користувачів в групи за спільними інтересами, та використання цих груп для надання цільових пропозицій.

Предмет дослідження – методи кластеризації користувачів соціальних мереж на основі великої кількості даних про їх діяльність в даних сервісах.

Наукова новизна отриманих результатів полягає в модифікації методу кластеризації розповсюдження схожості, за рахунок додавання властивості нечіткості. Та застосуванні модифікованого методу кластеризації в моделі підготовки цільових пропозицій на основі вподобань користувачів, та діяльності користувачів у соціальних мережах.

Публікації.

Касянчик Д. О. Використання методу k-середніх для об'єднання користувачів за даними з соціальних мереж / Д.О. Касянчик / Тринадцята міжнародна науково-практична конференція (МОДС. – 2018). – м. Чернігів.: ЧНТУ, 25 - 29 червня 2018 р. – С. 155-158

Касянчик Д.О. Інформаційна технологія кластеризації користувачів для підготовки цільових пропозицій / Д.О. Касянчик / Всеукраїнська науково-практична конференція молодих вчених та студентів «Інформаційні системи та технології управління» (ІСТУ-2018) – м. Київ.: НТУУ «КПІ ім. Ігоря Сікорського», 29-30 грудня 2018 р. – С. 16-19.

КЛАСТЕРИЗАЦІЯ, РЕКОМЕНДАЦІЇ, СОЦІАЛЬНІ МЕРЕЖІ, ВЕЛИКІ ДАНІ, БЕКОФІС, ФРОНТОФІС

ABSTRACT

Master's dissertation: 94 pages, 21 pictures, 32 tables, 1 appendix, 30 sources.

Topicality. Every day each of us receives quite a lot of information. Received information not always is relevant, which necessitates the search for sources that will be interesting and trustworthy. Most people have jobs, and many have some hobbies. In both cases, a person may face a task in learning something new, a deepening knowledge in a particular field. Also, almost every day watching movies or reading books, or listening to music. And, in order to do these actions, you must find material that will be interesting to you. Often a person diving in the huge amount of material when looking for exactly what she needs. For example, we can find up to 10, and sometimes more books on one topic, but we just don't have enough time to read all of it.

In the described cases it is necessary to ask someone advice, and of course we do not always have such an opportunity. In such cases, the aid is coming a recommender system.

Most of recommender systems issue proposals based on previously viewed or based on similar materials. One more significant problem of such systems is the focus on just one topic. These systems often can not cope with the role of counselor, when a person is interested in several topics. Also, if a person would like to know what is interesting to her like-minded people, then a recommendation based on previously viewed materials may simply not fit her.

Therefore, it is expedient to develop a technology to unite like-minded people based on activities in social networks. Through the use of technology from data science, the system will be able to cluster users with similar interests, which will allow users to recommend that part of his range of interests.

Relationship with academic programs, plans, themes. Work performed at the Department of ASOIU at the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" within the topic "Methods and technologies of high performance computing and performing of big data". Governments register number 0117U000924.

The goal of this research is quality improving and boosting of preparation speed of target proposals by optimizing a method of clustering users. To achieve the goal must perform the following tasks:

- analyze the existing means of obtaining data on the activity of users from the following social networks: facebook, twitter;
- collect data about user preferences and their activities in the facebook social network.
- choose the most suitable method for the preparation of targeted offers.
- select and modify the clustering method for its best use in the system being developed.
- implement the algorithm of preparing the target proposals based on the selected methods.

Object of research – the process of combining groups of people with common interests, and use these groups to provide targeted offers.

Subject of research – methods of clustering users of social networks based on large amounts of data on their activities in these services.

The scientific novelty of the results is to modify the clustering method Affinity Propagation, to the fuzzy Affinity Propagation. And using this modified clustering method in the model of preparation targeted proposals based on certain activities of users in social networks, and based on their preferences.

Published works.

Kasianchyk D. O. Using k-means for combining data from users of social networks [Electronic resource] / D. O. Kasianchyk / Thirteenth international scientific and practical conference (MODS. – 2018). – Chernihiv: ChNTU, June 25 – 29, 2018. – pp. 155-158.

Kasianchyk D.O. Information technology of user clustering for preparation of target proposals [Electronic resource] / D. O. Kasianchyk / All-Ukrainian scientific and practical conference of young scientists and students "Information Systems and Management Technologies" (ISMT – 2018). – Kyiv: NTUU "Igor Sikorsky Kyiv Polytechnic Institute", December 29-30, 2018. – pp. 16-19.

CLUSTERING, RECOMMENDATIONS, SOCIAL NETWORKS, BIG DATA,
BACK-OFFICE, FRONT-OFFICE

ЗМІСТ

ВСТУП.....	11
1 ПРОЕКТНІ РІШЕННЯ З РОЗРОБКИ СИСТЕМИ «Інформаційна технологія кластеризацій користувачів для підготовки цільових рішень»	14
1.1 Опис бізнес-процесів.....	14
1.1.1 Опис предметної області.....	14
1.1.2 Опис процесу діяльності.....	16
1.1.3 Актори та функції.....	17
1.1.4 Структура бізнес-процесів.....	21
1.2 Опис постановки задачі.....	25
1.2.1 Призначення розробки.....	25
1.2.2 Цілі та задачі системи	25
1.3 Рішення з інформаційного забезпечення	25
1.3.1 Вхідні дані	25
1.3.2 Вихідні дані	26
1.3.3 Структура бази даних.....	26
Висновок до розділу	32
2 МОДЕЛІ ТА МЕТОДИ КЛАСТЕРИЗАЦІЇ КОРИСТУВАЧІВ ТА ПІДГОТОВКИ ЦІЛЬОВИХ ПРОПОЗИЦІЙ.....	33
2.1 Змістовна постановка задачі.....	33
2.2 Математична модель	35
2.3 Огляд методів розв’язання.....	37
2.3.1 Огляд методів підготовки цільових пропозицій	37
2.3.2 Огляд методів кластеризації.....	38
2.4 Модифікація методу розв’язання задачі	39
2.5 Розробка алгоритму розв’язання.....	40
2.6 Результати досліджень ефективності методу	41
Висновок до розділу	47
3 ОПИС ПРОГРАМНОГО ТА ТЕХНІЧНОГО ЗАБЕЗПЕЧЕННЯ	48
3.1 Засоби розробки	48
3.2 Архітектура програмного забезпечення.....	52
3.2.1 Діаграма класів	52
3.2.2 Діаграма послідовностей	52
3.2.3 Архітектура Веб частини.....	55

	10
3.3 Інструкція адміністратора.....	56
3.3.1 Призначення підсистеми back-office.	56
3.3.2 Основні дій адміністратора	56
3.4 Інструкція редактора	57
3.4.1 Призначення підсистеми back-office	57
3.4.2 Основні дій редактора.....	57
3.5 Інструкція користувача	59
3.5.1 Призначення підсистеми front-office.....	59
3.5.2 Основні дій користувача.....	59
3.6 Опис технічного забезпечення	62
Висновок до розділу	62
4 РОЗРОБКА СТАРТАП-ПРОЕКТУ	64
4.1 Опис ідеї проекту (товару, послуги, технології)	64
4.2 Технологічний аудит ідеї проекту	68
4.3 Аналіз ринкових можливостей запуску стартап-проекту	69
4.4 Розроблення ринкової стратегії проекту	75
4.5 Розроблення маркетингової програми стартап-проекту.....	77
Висновок до розділу	79
ВИСНОВКИ.....	81
ПЕРЕЛІК ПОСИЛАНЬ	83
ДОДАТОК А Графічний матеріал.....	87
ER діаграма.....	88
Математична модель методу підготовки цільових пропозицій користувачам з використанням нечіткої кластеризації	89
Схема алгоритму підготовки цільових пропозицій користувачам з використанням нечіткої кластеризації	90
Схема алгоритму роботи модифікованого методу кластеризації на основі методу поширення схожості.....	91
Результати експерименту.....	92
Структурна схема класів.....	93
Структурна схема розгортання	94

ВСТУП

Об'єднана певним чином група людей має більше можливостей, аніж одинока людина для досягнення якоїсь мети. Об'єднання сприяє пошуку ідей, обмінну знаннями, підтримці як фізичній так і моральній. Люди протягом всього існування об'єднувалися за родинними ознаками, за інтересами, за професійною діяльністю, тощо. Дана форма соціальної діяльності людини збереглася і нині, проте якщо раніше для отримання переваг, потрібно було зазвичай брати безпосередню участь у діяльності групи, то нині розвиток технологій значно це спростив. Окрім напевно найвідомішого засобу комунікації - телефону, велику популярність мають соціальні мережі.

Соціальна мережа — веб-сайт або інший сервіс у мережі Інтернет, який дозволяє користувачам створювати публічний або частково публічний профіль, формувати список друзів з інших користувачів, обмінюватися текстовими повідомленнями, зображеннями, та іншими візуальними матеріалами а також переглядати профілі інших користувачів. Окрім, даного визначення соціальні мережі також є дуже потужною системою для обміну інформацією з користувачами і в першу чергу з користувачами, які відносяться до певної соціальної групи об'єднаної набором факторів. А це означає, що користувач може бути одночасно членом певної групи, та отримувати від цього досить великі переваги.

Також виникнення соціальних мереж в всесвітній павутині Інтернет дало змогу науковцям різних фахів аналізувати поведінку найрізноманітніших об'єднань людей.

Тому можемо зробити висновок, що соціальні мережі розширяють можливості з пошуку, та поширення інформації для користувачів зацікавлених в цій інформації. Також це дає змогу науковцям, а відповідно далі і виробникам товарів, авторам книг, статей і тому подібного аналізувати і отримувати інформацію про інтереси та потреби користувачів.

Проте можуть виникати і проблеми. Існує досить велика кількість випадків, коли виникає необхідність спитати в когось поради, і звісно не завжди в нас є така

можливість. В таких випадках на допомогу, приходять саме рекомендаційні системи.

Існуючі рекомендаційні систем можуть мати такі недоліки, як видача пропозиції лише на основі раніше переглянутих або на основі схожих матеріалів, а також вони можуть бути спрямованими лише на якусь одну тему. Тому вони часто не справляються з роллю порадирика, коли особу цікавлять декілька тем. Також у випадку, якщо людина хотіла б дізнатися, що цікаво її однодумцям, то рекомендація на основі раніше переглянутих матеріалів може їй просто не підійти

Це зумовлює необхідність розробки технології для об'єднання однодумців на основі діяльності в соціальних мережах. Адже рекомендації будуть відповідно до того, що цікаве об'єднанню людей за певним фактором, до якого входить і даний користувач.

Існує напевне, необмежена кількість тем, за якими можуть об'єднуватися користувачі:

- розваги, які в свою чергу включають;
- професійна діяльність, яка в свою чергу ділиться на професії;
- наукова діяльність, яка може ділитися на сфери;
- навчання, яке може ділитися на напрямки;
- хобі, яке напевне має необмежену кількість видів;
- геолокацією;
- тощо.

Дане дослідження виконується для об'єднання користувачів соціальних мереж за максимально схожою поведінкою, при цьому користувачі не обов'язково є друзями. Що буде використано для підготовки пропозиції користувачам відповідно до того, що є цікавим даній групі людей.

Система повинна бути спроектована так, щоб максимально зручно отримувати дані з соціальних мереж, збирати дані в межах системи, аналізувати їх, кластеризувати користувачів відповідно до даних про поведінку та вподобання, підготовляти цільові пропозиції користувачам на основі отриманих кластерів. Також

система має передбачати зручний інтерфейс для наповнення даними адміністратором, а також для отримання рекомендації звичайним користувачем.

Також, далі наведено проблеми які можуть виникнути при розробці даної інформаційної системи:

Можуть виникнути проблему зі збором інформації з соціальних мереж. Найпростішим методом вирішення може бути запит до користувача на обробку його персональних даних.

Також, проблема може бути в пошуку необхідних даних в інформації, збереженні та обробці цих даних. Саме це обумовлює використання науки про дані.

Отже, система має надавати можливість кластеризувати користувачів зі схожими інтересами та надавати їм цільові пропозиції, а також інтерфейс для отримання та перегляду цих пропозицій. При цьому система має враховувати проблеми пов'язані з обробкою персональних даних, а також зберіганням та обробкою великих масивів даних.

1 ПРОЕКТНІ РІШЕННЯ З РОЗРОБКИ СИСТЕМИ

«Інформаційна технологія кластеризацій користувачів для підготовки цільових рішень»

1.1 Опис бізнес-процесів

1.1.1 Опис предметної області

Розвиток технологій надав можливість без особливих проблем зберігати та оброблювати інформацію. І відповідно людство скористалося цією можливістю і створило необмежену кількість інформації, до великої частини з якої будь хто з нас має доступ. І особливо важливу ролі в обміні інформацією відіграє глобальна мережа Інтернет. Ми можемо користуватися онлайн бібліотеками та довідниками щоб дізнатися щось нове, також можемо дізнатися останні новини в будь яку секунду, просто скориставшись відповідними ресурсами, також ми можемо ділитися моментами свого життя, думками, переживаннями, ідеями з близькими і не тільки людьми, які знаходяться на великій відстані від нас, і для цих цілей створено велике розмаїття соціальних мереж, і це далеко не весь перелік інформації яку, ми можемо отримати з мережі Інтернет.

Проте, через об'єми інформації, які існують на сьогодні, є велика можливість просто потонути в інформаційному потоці. До прикладу: ми можемо отримувати новини з різних ресурсів, і кожен ресурс буде подавати її по-своєму; ми можемо шукати книгу, яку почитати, або фільм, який подивитися, а їх написано і знято така кількість, що не вистачить і життя щоб зі всіма ознайомитися. Саме тому розробляють засоби, які дозволяють проаналізувати великий об'єм даних і виокремити з них корисну інформацію. Ці засоби в свою чергу використовуються в сервісах, які надають звичайним користувача простий та зрозумілий інтерфейс для отримання саме корисної інформації.

Проблему з новинами вирішують сервіси по виявленню так званих “fake news”, проблему з підбором книги, фільму, тощо вирішують рекомендаційні системи. Рекомендаційні системи на основі певної діяльності користувача

допомагають отримати перелік об'єктів, що можуть бути йому цікаві, з мільйонів об'єктів. Даний сервіс “допомоги” користувачу можна значно покращити, використавши інформацію про поведінку користувача в соціальній мережі, адже більшість людей проводить досить багато часу в соціальних мережах. Мережі в свою чергу дають достатньо інформації, яку можна використати для підготовки цільових пропозиції користувачу.

Оскільки застосунок, що розробляється в межах цієї роботи, використовується для рекомендації фільмів, музики та книг, то розглянемо декілька ресурсів з максимально подібним функціоналом:

- IMDB (фільми, телесеріали);
- Netflix (фільми, телесеріали);
- Google Play Music, Apple Music (музика);
- YouTube (музика, відео блоги, телесеріали, телепередачі);
- Rozetka (товари).

IMDB є досить великою та зручною базою даних про фільми та телесеріали. Рекомендаційна система підбирає максимально схожі фільми, до того який користувач переглядає. Проте даний сервіс використовується лише для кінематографа.

Netflix схожий на IMDB тим, що використовується для фільмів та телесеріалів та телепередач. Дана платформа по факту є прокатом мультимедійних матеріалів. Даний сервіс напевне має найкращу систему рекомендацій, адже він враховує, що саме може зацікавити конкретного користувача жанр, актор, гама кольорів, режисер, тощо і відповідного до цього, рекомендації навіть одного і того ж фільму відбуваються з різними трейлерами або постерами. Проте цей сервіс також використовується лише для кінематографа.

Такі сервіси як YouTube, Google Play Music, Apple Music використовують досить потужні алгоритми для надання користувачу рекомендації на основі його діяльності, як в даних сервісах так і поза ними, проте основною функцією сервісів є надання можливості прослуховувати музику, та переглядати відеозаписи, серед яких дуже мала кількість фільмів.

Rozetka – український інтернет магазин, який враховує переглянуті товари користувачем, та що дивилися люди, які також це переглядали. Побічний спосіб використання даного ресурсу, може бути підбір книг користувачем.

Отже, кожен з цих сервісів має ряд переваг, а також певні недоліки, які обумовлені тим, що вони націлені на якусь одну сферу. Тобто, вони не є в першу чергу системою підбору персональних пропозиції для різних ресурсів, та матеріалів. Це дає змогу розробити систему, яка зможе зайняти свою нішу, як система підбору цільових пропозиції матеріалів різних типів.

1.1.2 Опис процесу діяльності

Соціальними мережами в Інтернеті є застосунки, які надають змогу користувачу створювати певний профіль, основними елементами якого зазвичай є:

- повне ім'я;
- фото або певне зображення;
- дата народження;
- інтереси;
- певні вподобання.

Ще одним важливим елементом є місце де користувач може здійснювати певну діяльність, яка буде видна іншим особам. Зазвичай таким місце є «стіна» де користувач розміщує певні публікації, які можуть бути текстовими, зображеннями, відеозаписами або комбінованими. Інші ж користувачі мають змогу, якось реагувати на ці публікації. Окрім цього важливим в соціальній мережі є функція додавання людей в друзі, що дає змогу бачити публікації в першу чергу людей, які є в цьому списку друзів, також виконувати дії, які можуть бути заблоковані для всіх інших користувачів, наприклад надсилати приватні повідомлення.

Саме цей великий спектр соціальної діяльності надасть змогу порівняти користувача, з іншими, та отримати інформацію про його вподобання.

Системи ж надання рекомендації сьогодні активно використовуються в найрізноманітніших сервісах. Це можуть бути рекомендації товарів в інтернет

магазинах, рекомендації друзів в соціальних мережах, рекомендації фільмів або музики а спеціалізованих сервісах, тощо.

В даній роботі описані автоматизація, підготовки даних отриманих з соціальної мережі facebook, об'єднання користувачів в кластери та підготовка користувачам цільових пропозиції.

1.1.3 Актори та функції

Акторами системи є:

- користувач системи;
- редактор;
- адміністратор.

Наведемо опис дій кожного з акторів.

Адміністратори можуть виконувати наступні функції:

- управляти користувачами;
- модерувати діяльність користувачів (видаляти неприпустимі коментарі, банити користувачів)

Редактори можуть виконувати наступні функції:

- створювати, редагувати та видаляти інформацію про матеріали (фільми, телепередачі, музика, книги);
- писати власні відгуки про нові матеріали;
- переглядати інформацію про кластери користувачів;
- додавати до матеріалів рекламу.

Редактори можуть виконувати наступні функції:

- отримувати та переглядати рекомендацій;
- шукати матеріали;
- переглядати матеріали;
- оцінювати матеріали;
- писати відгуки;
- перегляд та редагування особистої інформації та інформації, що використовується для надання рекомендацій.

В таблиці 1.1 наведені варіанти використання для всіх акторів системи.

Таблиця 1.1 – Опис дії варіантів використання

Актор	Назва варіанту	Опис	Пріоритет
Адміністратор	Управління користувачами	Адміністратор керує правами користувача в системі. За необхідності видаляє профіль користувача.	Високий
	Модерація діяльності користувачів	Адміністратор переглядає коментарі, залишені користувачами, та видаляє у випадку наявності неприпустимого вмісту. Якщо ситуація повторюється, блокує користувачу доступ до ресурсу на певний час.	Низький
Редактор	Створення, редагування та видалення інформації про матеріали	Редактор додає та управляє інформацією про фільми, телепередачі, музику та книги	Високий
	Написання власних відгуків про нові матеріали	Редактор описує власну думку про певний матеріал та прикріплює це, як відгук до публікації	Середній

Продовження таблиці 1.1

Актор	Назва варіанту	Опис	Пріоритет
	Перегляд інформації про кластери користувачів	Редактор може переглядати інформацію про кластери, такі як розміри, узагальнений портрет та ставлення користувачів кластеру до певного матеріалу	Високий
	Додавання до матеріалів реклами	На основі попереднього варіанту, редактор може допомогти рекламодавцю підготувати рекламу, та додати її до певного матеріалу	Високий
Користувач	Отримувати та переглядати рекомендацій	Користувач отримує список матеріалів з коротким описом, які його можуть зацікавити.	Високий
	Шукати матеріали	Користувач може переглянути весь список ресурсів, також може використати фільтри або пошук, щоб зменшити кількість матеріалів, що переглядаються	Високий
	Перегляд інформації про кластери користувачів	Редактор може переглядати інформацію про кластери, такі як розміри, узагальнений портрет та ставлення користувачів кластеру до певного матеріалу	Високий

Продовження таблиці 1.1

Актор	Назва варіанту	Опис	Пріоритет
Користувач	Оцінювати матеріали	Користувач може поставити власну оцінку матеріалу, яка враховується при підборі фільмів і яка вплине на рейтинг фільму	Високий
	Писати відгуки	Користувач може написати розгорнутий відгук, якщо він вважає, що оцінки не достатньо.	Низький
	Перегляд та редагування особистої інформації та інформації, що використовується для надання рекомендацій	Користувач, може переглянути та змінити, видалити особисті дані. Також він може переглянути, які дані впливають на підготовку цільових пропозицій, та у випадку необхідності внести зміни.	Середній

Відповідно визначених варіантів використання побудовано модель варіантів використання, яка наведена на рисунку 1.1. Дані варіанти використання визначені для акторів: адміністратора, редактора та користувача.

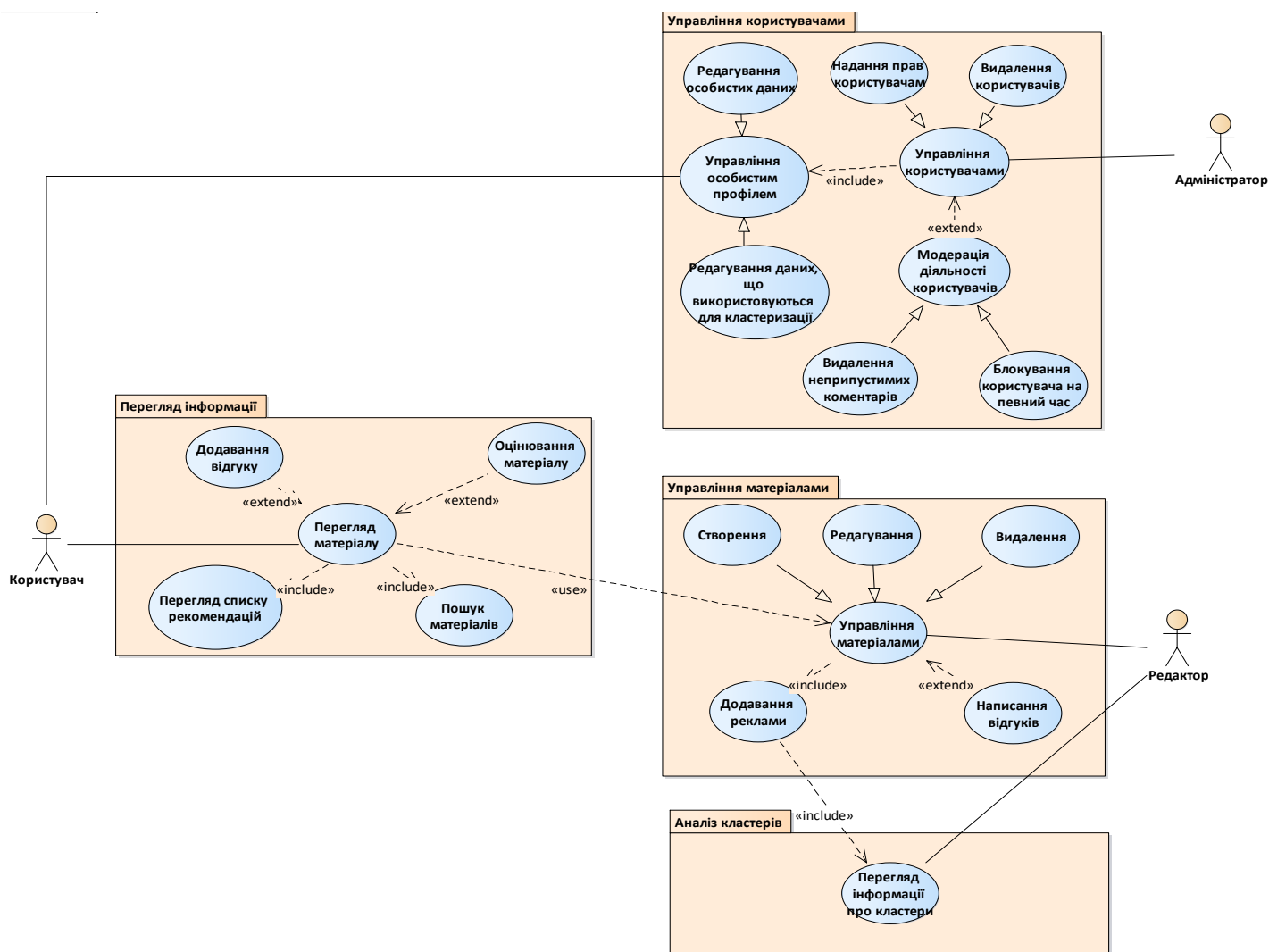


Рисунок 1.1 – Схема структурна варіантів використання

1.1.4 Структура бізнес-процесів

Дії, які має виконати адміністратор для управління користувачами, та модерації їх діяльності в системі розглянуто на рисунку 1.2.

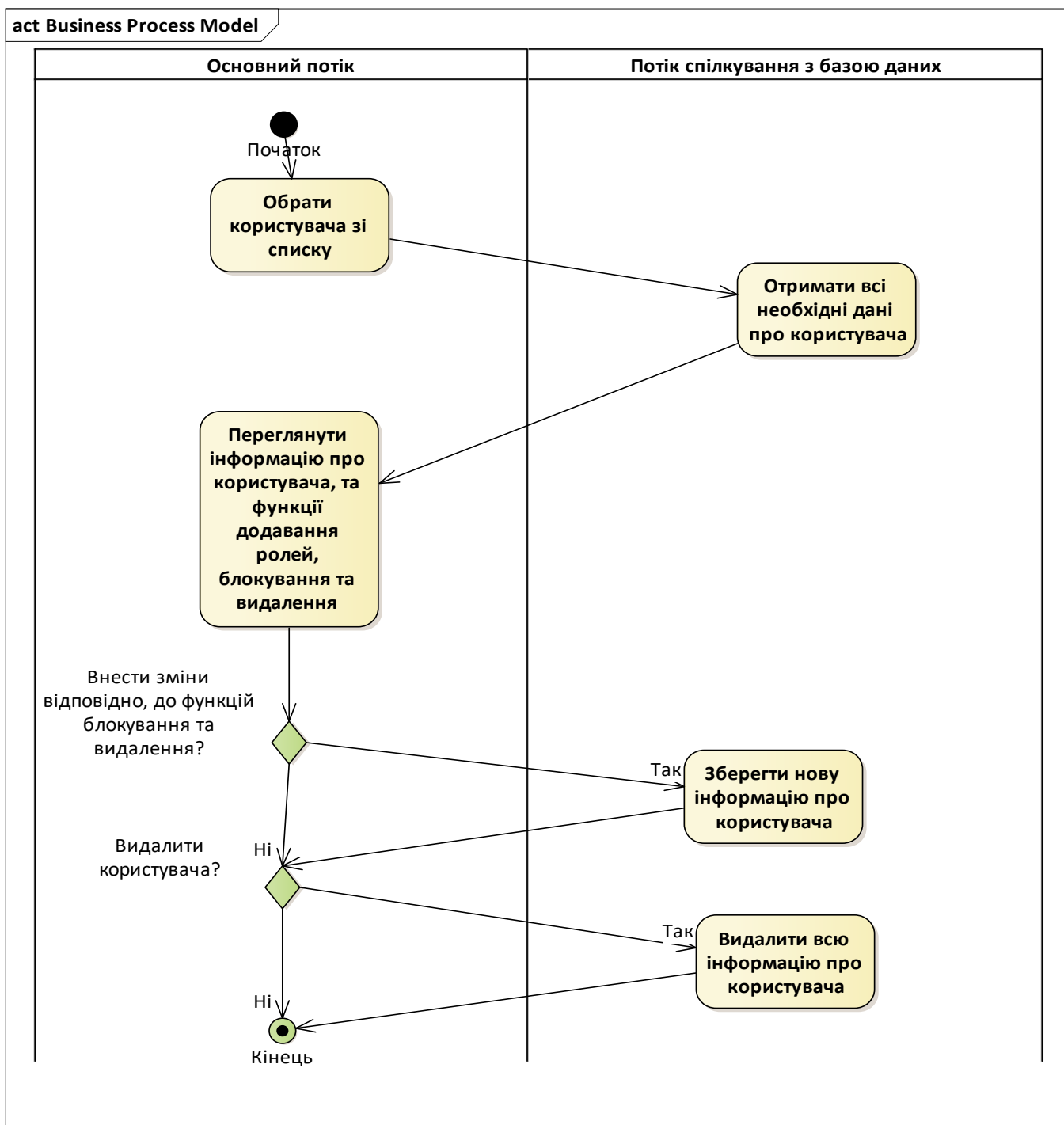


Рисунок 1.2– Схема структурна діяльності адміністратора

Розглянемо дії, які виконує редактор для наповнення системи матеріалами та додання реклами від партнерів (рисунок 1.3).

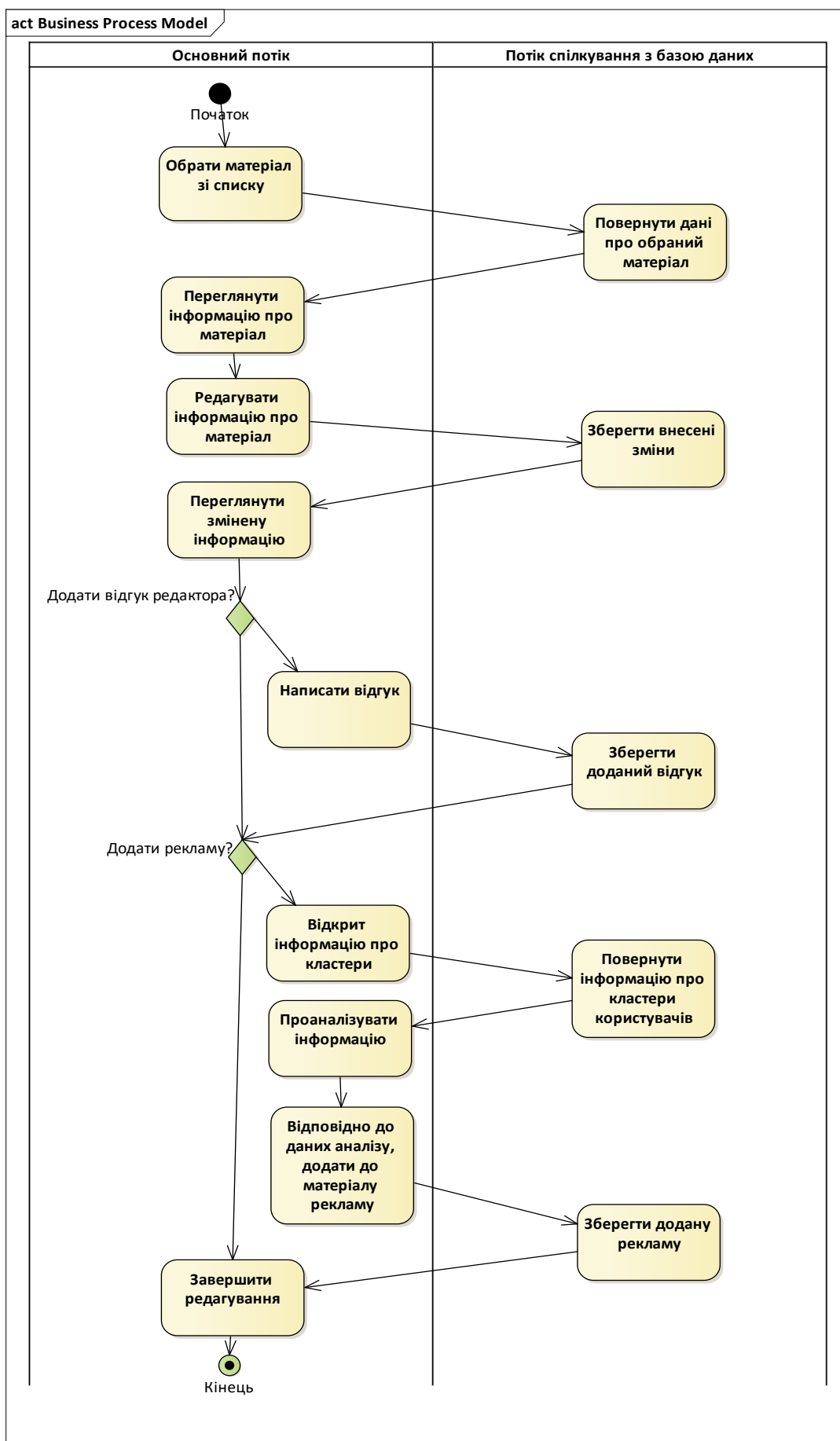


Рисунок 1.3 – Схема структурна діяльності редактора.

Розглянемо дії, які виконують в процесі діяльності користувача в системі (рисунок 1.4). Опишемо в схемі діяльність користувача яка пов'язана зі входом в систему та з пошуком, підбором, та переглядом матеріалів.

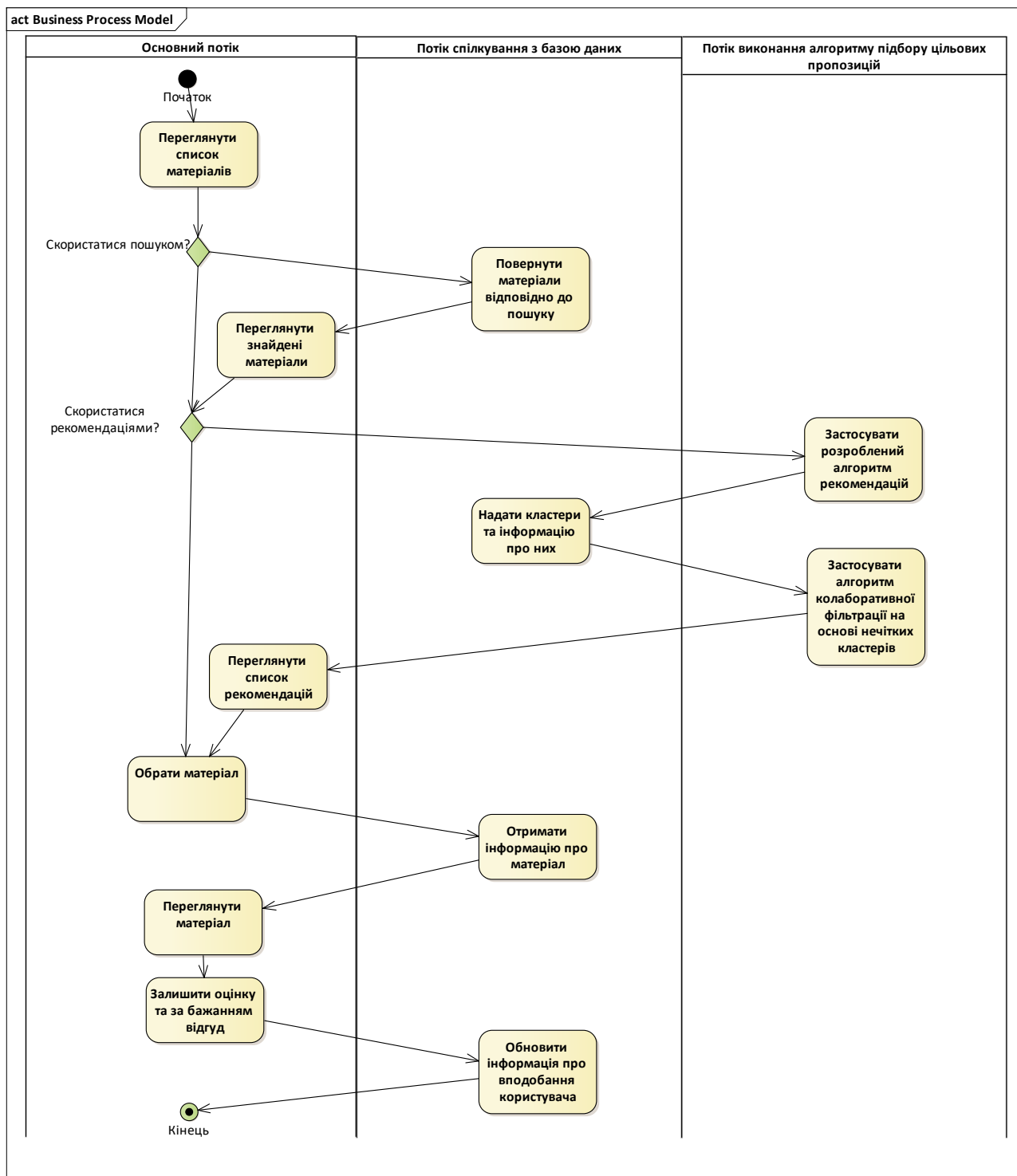


Рисунок 1.4 – Схема структурна діяльності користувача.

1.2 Опис постановки задачі

1.2.1 Призначення розробки

Інформаційна система надання цільових пропозицій користувачу, на основі нечіткої кластеризації:

- отримання даних про поведінку користувачів в соціальних мережах;
- об'єднання найбільш схожих користувачів на основі їх діяльності, за допомогою нечіткої кластеризації;
- надання цільових пропозицій на основі отриманих кластерів, та належності користувача до кластеру.

1.2.2 Цілі та задачі системи

Основні цілі системи:

- а) наповнення системи матеріалами:
 - 1) надання прав користувачам адміністратором;
 - 2) внесення та редагування матеріалів редактором.
- б) надання рекомендації користувачу:
 - 1) кластеризація користувачів;
 - 2) надання рекомендації користувачам, на основі належності до кластерів;
 - 3) врахування подальшої діяльності користувача системою.

1.3 Рішення з інформаційного забезпечення

1.3.1 Вхідні дані

Загалом вхідними даними є профілі користувачів системи з мережі facebook, а також оцінки які користувач поставив матеріалам в системі.

При першому вході користувача в систему враховується лише його профіль з соціальної мережі facebook.

1.3.2 Вихідні дані

Вихідними даними системи є сформовані кластери користувачів на основі їхньої поведінкової моделі та список рекомендації кожному користувачу на основі його належності до кластеру.

1.3.3 Структура бази даних

Для реалізації системи використовується реляційна база даних, тому зобразимо її за допомогою ER-діаграми та наведемо опис структури бази даних таблиці 1.2. У кожній сутності наявний атрибут id, який є первинним ключем.

Таблиця 1.2 – Опис структури реляційної бази даних

Назва таблиці	Назва атрибута	Таблиця на яку посилається	Тип даних	Детальна інформація
User	username		Char	нікнейм користувача
	first_name		Char	ім'я користувача
	last_name		Char	прізвище користувача
	email		Char	електронна пошта
	is_staff		Boolean	статус персоналу (False – звичайний користувач, True – адміністратор)
	is_superuser		Boolean	статус адміністратора з усіма правами (0 – користувач не має прав, 1 – має)
	is_active		Boolean	вказує чи активний профіль користувача

Продовження таблиці 1.2

Назва таблиці	Назва атрибута	Таблиця на яку посилається	Тип даних	Детальна інформація
User	date_joined		DateTime	дата створення аканту користувача
	last_login		DateTime	дата останнього входу в систему
FbInfo	home_address		Char	Домашня адреса в текстовому вигляді (країна, місто, вулиця)
	home_address_lat		Char	Широта домашньої адреси
	home_address_lng		Char	Довгота домашньої адреси
	location		Char	Поточне місце знаходження користувача в текстовому вигляді (країна, місто, вулиця)
	location_lat		Char	Широта адреси поточного місця знаходження
	location_lng		Char	Довгота адреси поточного місця знаходження
	age		Integer	Вік користувача
	gender		Char	Стать користувача (male, female)
	friends_count		Integer	Кількість друзів у facebook
	user	User	ForeignKey	Посилання на користувача

Продовження таблиці 1.2

Назва таблиці	Назва атрибута	Таблиця на яку посилається	Тип даних	Детальна інформація
Group	name		Char	Назва групи, до якої може належати користувач (адміністратор, редактор)
UserGroup	user	User	ForeignKey	Посилання на користувача багато до багатьох
	group	Group	ForeignKey	Посилання на групу у відношенні багато до багатьох
Permission	name		Char	Повна назва права користувача
	content_type	ContentType	ForeignKey	Посилання на вбудовану у фреймворку таблицю, яка вказує на певний модуль в додатку
	codename		Char	Короткий зміст колонки name для використання в програмному коді (для прикладу: "can_edit")
GroupPermission	group	Group	ForeignKey	Посилання на групу у відношенні багато до багатьох
	permission	Permission	ForeignKey	Посилання на дозвіл у відношенні багато до багатьох
Material	title_original		Char	Повна назва матеріалу, мовою оригіналу

Продовження таблиці 1.2

Назва таблиці	Назва атрибута	Таблиця на яку посилається	Тип даних	Детальна інформація
Material	decsription		Char	Опис матеріалу
	technical_description		Char	Технічний опис матеріалу, для кожного типу може бути різним (наприклад для відео – якість зображення, для книги – кількість сторінок, тощо)
	material_type	MaterialType	ForeignKey	Посилання на тип матеріалу
	average_mark		Float	Середня оцінка матеріалу, з усіх виставлених користувачами
	redactor_mark		Integer	Оцінка, виставлена редакторським складом
	image		Image	Зображення, яке відповідає даному матеріалу (наприклад, обкладинка книги, або музичного альбому, тощо)
MaterialTitleTranslate	material	Material	ForeignKey	Посилання на матеріал
	language		Char	Мова, якою наводиться переклад
	translate		Char	Власне переклад назви матеріалу

Продовження таблиці 1.2

Назва таблиці	Назва атрибута	Таблиця на яку посилається	Тип даних	Детальна інформація
MaterialType	title		Char	Назва типу матеріалу (фільм, книга, телепередача, музика)
	description		Char	Опис типу
	image		Image	Зображення, яке характеризує даний тип
	counter		Integer	Кількість матеріалів з даним типом
RedactorReview	material	Material	ForeignKey	Посилання на матеріал
	text		Char	Розгорнутий відгук редактора
	author		ForeignKey	Посилання на користувача, який додав відгук
Advertisement	picture		Image	Зображення реклами (логотип)
	text		Char	Опис реклами
	reference		Char	Посилання, за яким відбудеться перехід на сайт рекламодавця
	material	Material	ForeignKey	Зовнішній ключ на матеріал, до якого додається реклама
	author	User	ForeignKey	Посилання на користувача (редактора), який додав рекламу

Назва таблиці	Назва атрибута	Таблиця на яку посилається	Тип даних	Детальна інформація
Advertisement	date_start		Date	Дата, з якої починається відображення реклами
	date_end		Date	Дата, коли закінчиться відображення реклами
UserMark	material	Material	ForeignKey	Посилання на матеріал
	user	User	ForeignKey	Посилання на користувача, який додав оцінку
	mark		Integer	Оцінка користувача певному матеріалу
	is_from_fb		Boolean	Значення, яке вказує, чи є дана оцінка отриманою з ФБ
UserComment	material	Material	ForeignKey	Посилання на матеріал
	user		ForeignKey	Посилання на користувача
	comment_text		Char	Відгук залишений користувачем
	creation_date		Date	Дата, коли відгук був створений

Схема реляційної бази даних у вигляді ER діаграми наведена у додатку А, плакат «ER діаграма».

Також в системі використовується NoSQL база даних для збереження навченої моделі кластеризації, структура якої може частково змінюватися. Зобразимо цю базу за допомогою ER-діаграми та наведемо опис структури таблиці 1.3.

Таблиця 1.3 – – Опис структури NoSQL бази даних

Назва колекцій	Поле документа	Тип даних	Детальна інформація
Clustering	model	Binary	Навчена модель (екземпляр класу), яка зберігається у вигляді байт об'єктів, яку за необхідності можна перетворити назад у модель.

Схема NoSQL бази даних складається з одного елементу та наведена у вигляді ER діаграми на рисунку 1.5.



Рисунок 1.5 – ER діаграма NoSQL бази даних.

Висновок до розділу

В розділі описано предметне середовище, процес діяльності, акторів процесу діяльності та побудовано діаграму варіантів використання також побудовано діаграми діяльності для кожної з ролей.

В пункті опису постановки задачі сформульована мета створення системи, призначення її, та надано перелік задач, які треба реалізувати щоб досягнути мети.

В пункті рішень з інформаційної безпеки описано які дані приймає система та які отримуємо на виході. А також наведено ER – діаграму, яка показує, в якому вигляді дані зберігаються в системі.

Інформація в даному розділі покликана в загальному зрозуміти, як система працює, як система обробляє дані, та що очікується від неї. Це дає змогу розробляти моделі та методи відповідно до сформульованих в розділі вимог.

2 МОДЕЛІ ТА МЕТОДИ КЛАСТЕРИЗАЦІЇ КОРИСТУВАЧІВ ТА ПІДГОТОВКИ ЦІЛЮВИХ ПРОПОЗИЦІЙ

2.1 Змістовна постановка задачі

Заданий вхідний набір векторів користувачів, вектори враховують діяльність користувача в межах системи, що розробляється та діяльність користувача в соціальних мережах, в даному випадку лише в соц. мережі Facebook. Заданий набір матеріалів певного типу: фільми, музика, телепередачі, книги. Кожен з цих матеріалів має певний рейтинг, а також кожен користувач може висловити свою думку про нього у вигляді оцінки та коментаря (в даному випадку важлива лише оцінка, так як для перетворення коментарів в числові значення потрібно розробляти додаткові модифікації).

Необхідно на основі поведінки особи в межах сервісу, і на основі даних отриманих поза сервісом підготувати цільові пропозиції, які будуть задовільними для нього. Основна ціль даного сервісу зі сторони бізнесу: збільшити аудиторію, також збільшити час перебування користувачів на сайті, покращити надання рекламних пропозицій, дані зміни покликані в свою чергу збільшити прибуток бізнесу. Ціль зі сторони користувача: знизити час на пошук необхідного матеріалу, покращити досвід користування сайтом.

Завдяки тому, що є дані з соціальної мережі можна уже на початковому етапі надавати користувачеві певні рекомендації, а також коли навіть буде здійснена достатня кількість дій в системі, інформація з соціальної мережі повинна допомогти точніше надавати рекомендації [1]. Адже буде використовуватися більш повний профіль користувача.

Проблемами даної системи є перетворення даних у числові вектори та досягнення достатньої швидкості виконання алгоритму підготовки цільових пропозицій, також проблемою може бути локалізація назв матеріалів.

Для досягнення достатньої швидкості можна застосувати сучасні засоби розробки та машинного навчання, які дають можливість здійснювати «навчання»

моделі, не кожен раз за потребою користувача, а припустимо один раз на день, щодня з 3 до 4 години ночі (коли користувачі найменш активні).

Проблема локалізацією також вирішується сучасними засобами розробки, в даному випадку за допомогою пошукового двигуна.

Також, наведемо опис вирішення проблеми підготовки даних.

Для надання кластеризації та надання рекомендацій потрібні дані двох видів:

- які описують користувача;
- які відображають вподобання користувача.

Для отримання даних, що описують особу скористаємося відкритим API.

Виділимо інформацію, з тієї яку можна отримати, що буде використовуватися:

- домашня адреса;
- поточна геолокація;
- вік;
- стать;
- кількість друзів;
- лайки, які користувач поставив сторінкам: книг, фільмів, музики,

телепередач.

Для отримання даних, що описують вподобання користувача будуть використані оцінки користувачем певних об'єктів (книга, фільм, музика або телепередача) в даній рекомендаційній системі. Значення оцінки можуть бути від 0 до 5.

Для «холодного старту» використовуються лайки з facebook. Тобто зіставляються назви сторінок, яким користувач поставив лайк та об'єктам у системі. У випадку збігу об'єкту виставляється автоматично оцінка «5», у випадку, коли об'єкт не збігається з жодним лайком значення оцінки дорівнює «0».

Побудувати вектор з усіх наявних об'єктів, що описує користувача, та застосувати його для кластеризації, навряд є можливим. Тому кластеризація буде здійснюватися для кожного типу об'єктів окремо. Також у вектор буде входити топ 100 об'єктів даного типу за кількістю оцінок користувачів в даній інформаційній системі, наприклад топ 100 книг, що зібрали найбільшу кількість оцінок

користувачів. Даний вибір зумовлено тим, що користувача на основі ставлення до найбільш популярних книг можна віднести до певної групи.

2.2 Математична модель

Основною задачею є надання найбільш релевантних цільових пропозицій користувачу. Сформулюємо математичну постановку задачі.

Дано:

- множина користувачів U ;
- кількість користувачів – n ;
- вектор користувача – u_i , де $i = \overline{1, n}$;
- вектор користувача з яким виконується порівняння – u' , де $i = \overline{1, n}$;
- кількість матеріалів – m ;
- конкретний матеріал – j , де $j = \overline{1, m}$.

Змінні:

$r_{u,j}$ - можлива оцінка матеріалу користувачем (передбачення);

$r'_{u,j}$ - реальна оцінка матеріалу;

$sim(u, u')$ – функція схожості між користувачами.

Цільова функція – максимізувати кількість релевантних передбачених оцінок матеріалу:

$$Z = \frac{1}{m} \sum_j^m 1 - \frac{|r_{u,j} - r'_{u,j}|}{\max(r'_{u,j}, r_{u,j})} \rightarrow \max, \quad (2.1)$$

де

$$r_{u,j} = \sum_{u' \in U} sim(u, u') * r'_{u',j} \quad (2.2)$$

Обмеження:

$$0 < r_{u,j} < 5, \quad (2.3)$$

$$0 < r'_{u,j} < 5, \quad (2.4)$$

$$0 < sim(u, u') < 1 \quad (2.5)$$

Для розв'язку задачі обрано метод колаборативної фільтрації (нижче описано чому), а для обчислення схожості користувачів обрано методи машинного навчання, адже в поєднанні з сучасними засобами розробки отримаємо приріст швидкості виконання, а також зможемо об'єднувати більшу кількість людей з більшою кількістю параметрів.

Тому опишемо ще математичну постановку задачі методу кластеризації.

Дано множину користувачів U , кожен з яких є вектором атрибутів. Потрібно знайти таку функцію F , яка розіб'є множину на кластери C , та вкаже на ймовірність належності об'єкта кластеру $\mu(i)$ [2,3].

$$U = \{u_1, u_2, \dots, u_n\}, \quad (2.6)$$

де n - кількість об'єктів (користувачів).

$$C = \{c_1, c_2, \dots, c_k, \dots, c_h\}, \quad (2.7)$$

де h - кількість кластерів, яка в даному випадку не є зарання відомою.

c_k - кластер, що містить в собі схожі об'єкти з множини U .

$$c_k = \{u_i, u_g \mid u_i \in U, u_g \in U \ \& \ d(u_i, u_g) < \sigma \}, \quad (2.8)$$

де σ - величина, яка визначає значення близькості для включення об'єкту до кластеру, $d(u_i, u_g)$ - відстань між векторами користувачів.

$$\mu_k(u_i) = d(u_i, u_k), \quad (2.9)$$

$$0 < \mu_k(u_i) < 1, \quad (2.10)$$

$$\sum_{k=0}^h \mu_k(u_i) = 1, \quad (2.11)$$

де $\mu_k(u_j)$ - міра схожості поточного користувача, до кластера k .

Дані обмеження визначають, що об'єкт може належати різним кластерам з різними ступенями приналежності і сума приналежності об'єкта всім кластерам дорівнює одиниці.

2.3 Огляд методів розв'язання

2.3.1 Огляд методів підготовки цільових пропозицій

Розглянемо два основні методи та їх різновиди рекомендаційних систем. Порівняння методів надано в таблиці 2.1[4-7].

Таблиця 2.1 – Порівняння методів підготовки цільових пропозицій.

	Вхідні дані	Як використовується	Проблеми
Фільтрація на основі змісту	Об'єкти які користувач оцінив	Оцінюється схожість за параметрами об'єктів	Велика залежність від предметної області, ігнорування досвіду інших користувачів
Колаборативна фільтрація на основі пам'яті	Оцінки виставлені користувачем; Оцінки об'єктів , виставлені іншими користувачами	Знаходиться схожість між користувачами на основі їх оцінок	Проблема «холодного старту», проблема «білих ворон»
Колаборативна фільтрація на основі моделі (кластеризації)	Діяльність користувачів в соціальних мережах. Оцінки об'єктів , виставлені іншими користувачами	Будуються кластери користувачів на основі їх векторів і поточному надаються рекомендації відповідно до оцінок всередині кластеру	Не враховуються вподобання
Гібридна колаборативна фільтрація	Оцінки виставлені користувачем; Оцінки об'єктів , виставлені іншими користувачами; Діяльність користувачів в соц. мережах.	Будуються кластери користувачів на основі їх векторів з врахуванням вподобань і поточному надаються рекомендації відповідно до оцінок всередині кластеру	Згадані вище проблеми відсутні

2.3.2 Огляд методів кластеризації

Кластерний аналіз походить з природничих наук, і остаточно став відомим в середині минулого століття. З розвитком комп'ютерної техніки, почався новий етап розвитку даного напрямку. Саме тому існує досить велика кількість алгоритмів кластеризації. Всі ці алгоритми об'єднуються в певні моделі. Основними є наступні моделі [8-11].:

- моделі зв'язності;
- центроїдні моделі;
- статистичні моделі, де кластери будуються ґрунтуючись на статистичних розподілах;
- моделі засновані на щільності;
- групові моделі. дані алгоритми просто описують групування об'єктів;
- графові моделі;
- нейронні моделі.

Порівнюємо методи, що належать до описаних моделей. Порівняння наведене в таблиці 2.2. [12-16]

Таблиця 2.2 – Порівняння методів кластеризації.

Назва	Обов'язкові параметри	Коли використовується	Метрики
К-середніх	Кількість кластерів	Плоска геометрія, не дуже багато кластерів	Відстань між точками
Поширення близькості	згладжування, шум	Багато кластерів, нерівномірні розміри кластерів, неплоска геометрія	Графова відстань (наприклад, граф найближчих сусідів)
Середній зсув	пропускна здатність	Багато кластерів, нерівномірні розміри кластерів, неплоска геометрія	Відстань між точками

Назва	Обов'язкові параметри	Коли використовується	Метрики
Спектральна кластеризація	Кількість кластерів	Кілька кластерів, неплоска геометрія	Графова відстань (наприклад, граф найближчих сусідів)
Ієрархічна кластеризація	Кількість кластерів	Багато кластерів, можлива зв'язність обмежень	Відстань між точками
Агломеративна кластеризація	Кількість кластерів, тип зв'язку, відстань	Багато кластерів, можлива зв'язність обмежень, не евклідова відстані	Будь-яка попарна відстань
DBSCAN	Розмір сусідства	Неплоска геометрія, невідомі розміри кластерів	Відстань між найближчими точками

Кожен з даних методів, має свої переваги і недоліки. Проте для даної роботи було обрано саме метод поширення близькості (Affinity Propagation).

Перевагою даного методу є те, що не потрібно зарання задавати центри класів, а також кількість кластерів. Також нам не важлива кількість елементів, і відповідно її не потрібно задавати.

2.4 Модифікація методу розв'язання задачі

В першу чергу внесемо зміни в алгоритм надання рекомендацій, де замість стандартного обчислення схожості та пошуку схожих користувачів будемо використовувати нечітку кластеризацію.

Також, потрібно модифікувати сам метод кластеризації об'єктів, шляхом перетворення його на не чіткий. Тобто замість стандартної відстані між об'єктами

буде враховуватися відсоток належності поточного об'єкта до певного іншого. Для цього потрібно замінити початкову матрицю відстаней, на матрицю схожості.

Кластери тепер необхідно використати, як узагальнених користувачів. Це дає змогу використати колаборативну фільтрацію засновану на сусідстві. Де функція схожості, виглядає наступним чином:

$$sim(u_i, u_k) = \mu_k(u_i), \quad (2.12)$$

де, u_j – поточний користувача,

u_k – узагальнений користувача, тобто кластер,

$\mu_k(u_i)$ – міра відношення поточного користувача, до кластеру.

Мінусом, є трохи більша складність алгоритму, відповідно довше виконання, проте це компенсується на етапі реалізації інформаційної системи. Тобто ми зможемо виконувати кластеризацію в бекграунді непомітно для користувача, і він не буде змушений чекати підготовки пропозицій.

Також перевагою є те, що враховуватися буде думка не лише «схожих» користувачів, а всіх користувачів системи, яка звісно ж домножується на коефіцієнт близькості.

Також вагомим мінусом, даної системи можна назвати сильну прив'язку до соціальної мереж. Проте сьогодні складно уявити людину, яка не користується останніми.

2.5 Розробка алгоритму розв'язання

Модифікований алгоритм кластеризації побудований на основі методу поширення схожості. Основною ціллю модифікацій є додавання властивості нечіткості, тобто визначення міри належності користувача кожному з кластерів. В даному розділі наведено покроковий алгоритм модифікованого методу.

Крок 1. ПЕРЕДАТИ в метод матрицю відстаней, що обраховані за один зі стандартних методів.

Крок 2. ПЕРЕТВОРИТИ матрицю відстаней, у матрицю схожості між об'єктами.

Крок 3. СТВОРИТИ матрицю належності об'єктів кластерам

Крок 4. ПЕРЕЙТИ до ітерації.

Крок 5. ОБЧИСЛИТИ матрицю «відповідальності».

Крок 6 . ОБЧИСЛИТИ матрицю «доступності».

Крок 7. ОБРАТИ центри та обрати значення

Крок 8. ЗАПИСАТИ значення з матриці схожості в матрицю належності для центрів, що змінилися або додалися.

Крок 9. ЯКЩО остання ітерація перейти до (9), ІНАКШЕ до (5).

Крок 10. ЗБЕРЕГТИ центри та матрицю схожості.

На цьому робота алгоритму завершується.

Алгоритм колаборативної фільтрації не наводиться, адже він змінився не суттєво. Зміна полягає лише в тому, що використовуються кластери, як узагальнені користувачі замість всієї множини користувачів.

2.6 Результати досліджень ефективності методу

Для порівняння алгоритмів використаємо наступні метрики: гомогенність, повнота, V-міра, силует. Нижче наведені формули[17,18] для обчислення обраних метрик та порівняння методів (таблиця 2.3).

$$h = 1 - \frac{H(C|K)}{H(C)}, \quad (2.13)$$

$$c = 1 - \frac{H(K|C)}{H(K)}, \quad (2.14)$$

$$v = 2 \frac{hc}{h + c}, \quad (2.15)$$

$$s = \frac{b - a}{\max(a, b)} \quad (2.16)$$

Гомогенність та повнота обчислюються за допомогою функції ентропії та умовної ентропії. Де К – результат кластеризації, С – відомий розподіл на кластери. Дані метрики лежать в межах від 0 до 1, і чим більше значення тим точніша кластеризація. Для врахування обох величин була введена V-міра.

Для визначення силуету нам непотрібний відомий розподіл на кластери. Де a – середня відстань від об'єкта до об'єктів в межах кластеру, b – середня відстань від об'єкта до об'єктів найближчого кластеру.

Оскільки, ще не відомі реальні розбиття датасету користувачів на кластери, то порівняння методів відбувається на основі двомірних масивів точок різних розмірів та форм. Дані методи порівняно за вказаними метриками, а також двомірність даних точок надає можливість їх візуалізувати. В таблиці 2.3 наведено порівняння методів для 100 точок, які мають довільне розміщення.

Таблиця 2.3 – Порівняння методів кластеризації для 100 об'єктів.

	Гомогенність	Повнота	V-міра	Силует	Час виконання, с
К-середніх	0.677	0.705	0.691	0.762	0.02
Спектральна	0.864	0.871	0.868	0.721	0.04
Агломеративна	0.584	0.935	0.719	0.517	0.01
DBSCAN	0.731	0.756	0.743	0.374	0.004
Поширення схожості	0.000	1.000	0.000	0.000	0.01
Модифікований метод поширення схожості	0.000	1.000	0.000	0.000	0.02

На рисунку 2.1 наведено візуалізоване порівняння методів для 100 точок розміщених у різних формах.

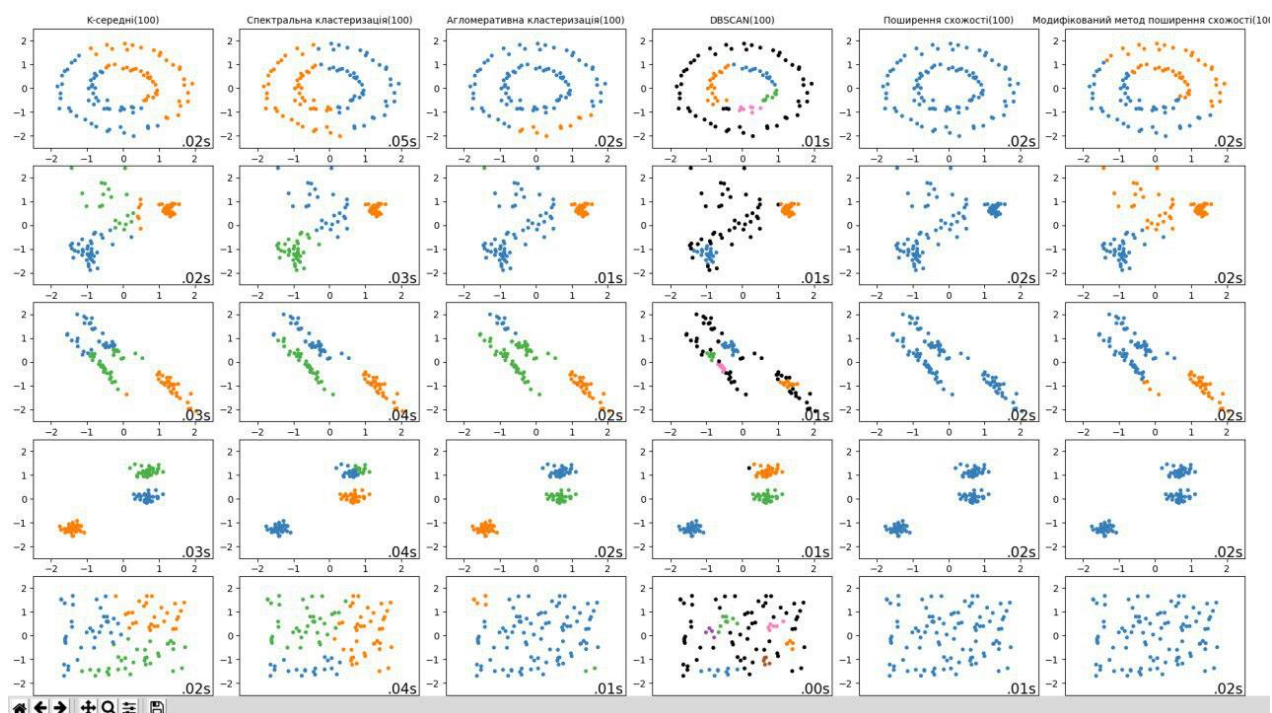


Рисунок 2.1 – Візуалізоване порівняння методів для 100 точок.

В таблиці 2.4 наведено порівняння методів для 500 точок в двовимірному просторі, які мають довільне розміщення.

Таблиця 2.4 – Порівняння методів кластеризації для 500 об'єктів.

	Гомогенність	Повнота	V-міра	Силует	Час виконання, с
К-середніх	0.749	0.762	0.755	0.781	0.04
Спектральна	0.795	0.804	0.800	0.765	0.10
Агломеративна	0.486	0.790	0.602	0.592	0.06
DBSCAN	0.907	0.693	0.786	0.350	0.004
Поширення схожості	0.463	0.757	0.575	0.804	0.37
Модифікований метод поширення схожості	0.813	0.698	0.751	0.779	0.37

На рисунку 2.2 наведено візуалізоване порівняння методів для 500 точок розміщених у різних формах.

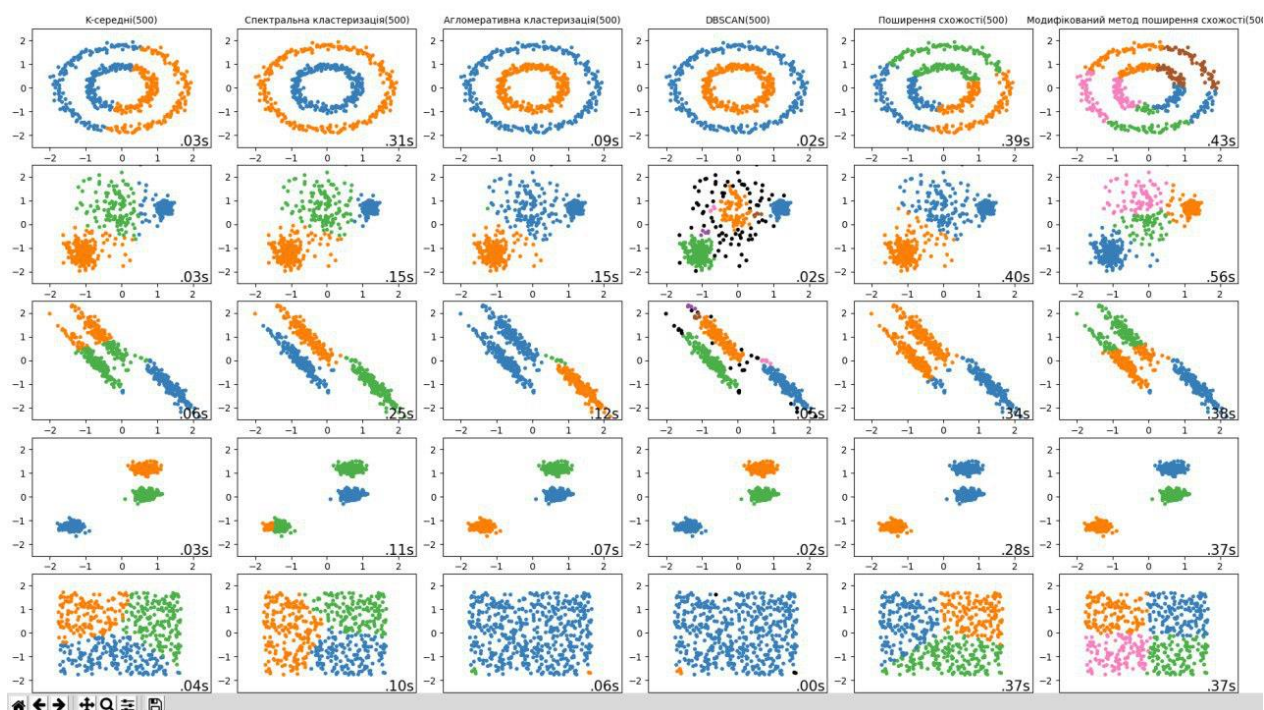


Рисунок 2.2 – Візуалізоване порівняння методів для 500 точок.

В таблиці 2.5 наведено порівняння методів для 1000 точок в двовимірному просторі, які мають довільне розміщення.

Таблиця 2.5 – Порівняння методів кластеризації для 1000 об'єктів.

	Гомогенність	Повнота	V-міра	Силует	Час виконання, с
К-середніх	0.802	0.810	0.806	0.762	0.10
Спектральна	0.925	0.925	0.925	0.718	0.14
Агломеративна	0.937	0.937	0.937	0.719	0.70
DBSCAN	0.928	0.779	0.847	0.643	0.01
Поширення схожості	0.795	0.803	0.799	0.763	0.82
Модифікований метод поширення схожості	0.853	0.721	0.781	0.772	1.25

На рисунку 2.3 наведено візуалізоване порівняння методів для 1000 точок розміщених у різних формах.

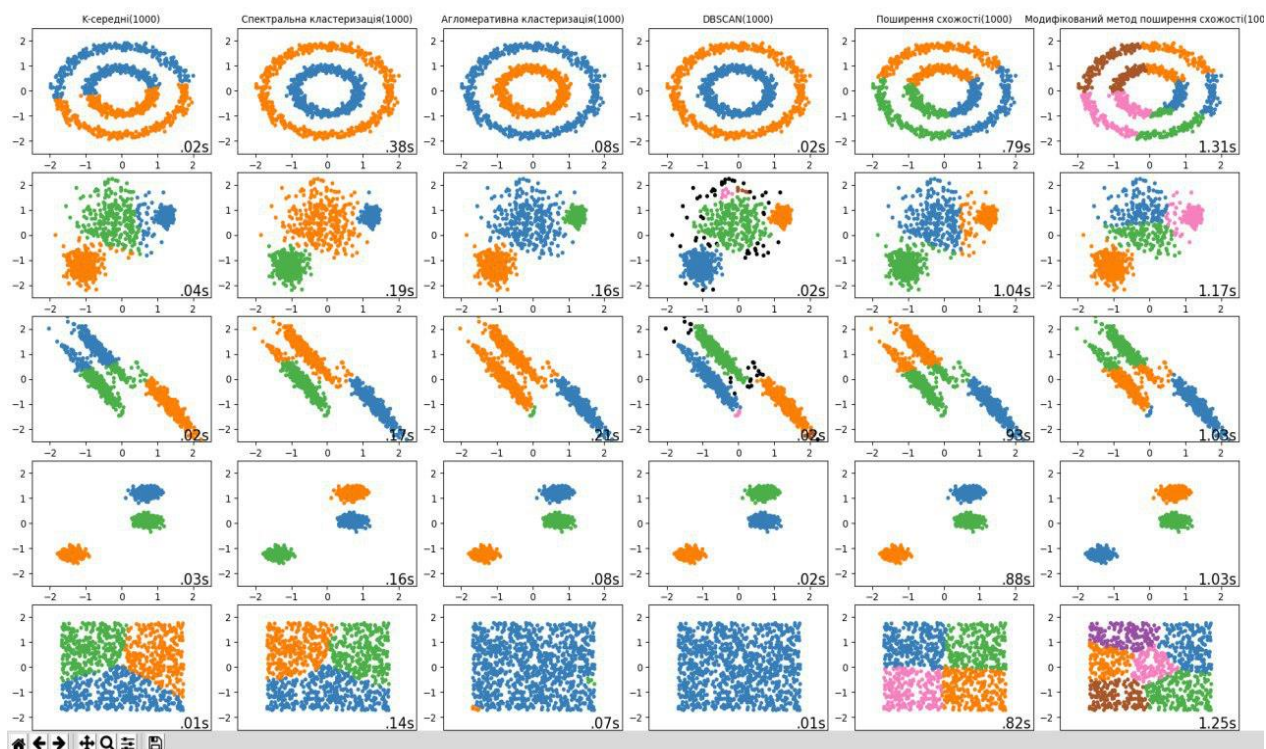


Рисунок 2.3 – Візуалізоване порівняння методів для 1000 точок.

В таблиці 2.6 наведено порівняння методів для 5000 точок в двовимірному просторі, які мають довільне розміщення.

Таблиця 2.6 – Порівняння методів кластеризації для 5000 об'єктів.

	Гомогенність	Повнота	V-міра	Силует	Час виконання, с
К-середніх	0.779	0.789	0.784	0.773	0.30
Спектральна	0.886	0.887	0.886	0.726	1.93
Агломеративна	0.891	0.892	0.892	0.714	0.38
DBSCAN	0.007	0.150	0.013	0.389	0.06
Поширення схожості	0.839	0.712	0.771	0.779	30.50
Модифікований метод поширення схожості	0.898	0.528	0.665	0.634	34.02

На рисунку 2.4 наведено візуалізоване порівняння методів для 5000 точок розміщених у різних формах.

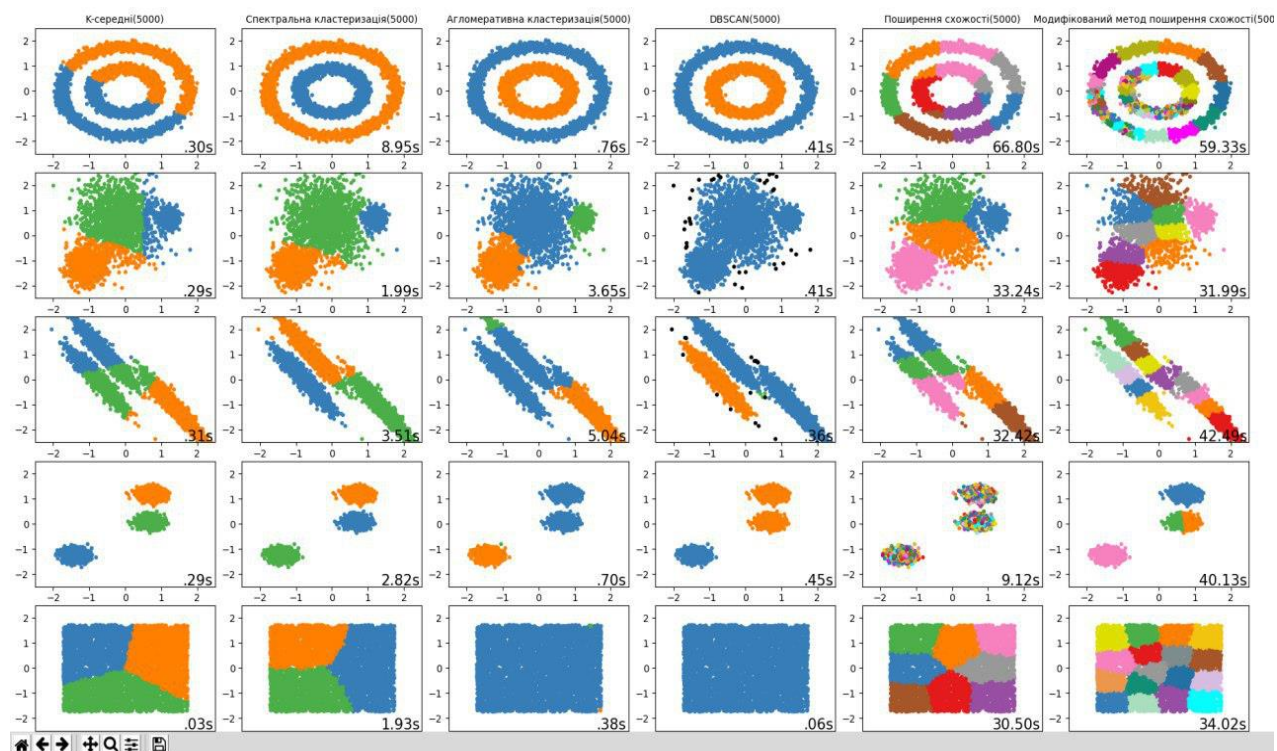


Рисунок 2.4 – Візуалізоване порівняння методів для 5000 точок.

За результатами наведеними у таблицях 2.3 – 2.6 побудовано графіки порівнянь, що наведено у додатку А, плакат «Результати експерименту».

Як бачимо, модифікований методу дає лише трохи кращі результати, ніж не модифікований, але дає гірші результати в часі. Проте як було описано раніше, дана проблема вирішується за допомогою сучасних засобів розробки. Також модифікований метод дає гірші результати, ніж інші методи за деякими характеристиками. Проте це компенсується тим, що не потрібно зарання знати кількість кластерів, їх розміри та центри. Як видно з візуалізованого порівняння методів, модифікований метод поширення схожості краще підходить у випадку тісного скупчення точок, що є його перевагою у випадку кластеризації користувачів соціальних мереж. Також завдяки цьому методи, отримано міру схожості користувача до узагальненого користувача (кластера), що буде використано в методі колаборативної фільтрації.

Висновок до розділу

В розділі надано змістовну та математичну постановку задачі надання цільових пропозицій користувачам на основі кластерів.

Оглянуто методи підготовки цільових пропозицій та обрано метод колаборативної фільтрації. Також оглянуто методи кластеризації користувачів та обрано метод поширення близькості. Оскільки даний метод не мав міри належності користувача до кластера, для підготовки цільових пропозицій, то його було модифіковано.

Модифікований метод показав не гірший результат ніж інші методи , що дозволило використати його, а саме властивість нечіткості цього модифікованого методу для підготовки цільових пропозицій.

Розробка цієї моделі дала змогу, реалізувати інформаційну систему яка надає цільові пропозиції кожному з користувачів.

3 ОПИС ПРОГРАМНОГО ТА ТЕХНІЧНОГО ЗАБЕЗПЕЧЕННЯ

3.1 Засоби розробки

Для розробки програмного продукту використано такі засоби для розробки бек-енд частини: Python та фреймворк Django; для фронт-енд частини: JavaScript та бібліотека jQuery, а також HTML5 та Bootstrap, для обробки даних: бібліотека для машинного навчання scikit-learn та пошуковий двигун Elasticsearch, а також база даних PostgreSQL та MongoDB.

Python - це проста у вивченні, і водночас потужна мова програмування. Вона має ефективні високорівневі структури даних та просту, але ефективну реалізацію об'єктно-орієнтованого програмування. Елегантний синтаксис Python і динамічна типізація, поряд із інтерпретованістю, роблять його ідеальною мовою сценаріїв та швидкого розроблення потужних додатків у багатьох областях на більшості існуючих платформ. Інтерпретатор Python і доволі широка стандартна бібліотека є вільно поширюваними для всіх основних платформ.

Інтерпретатор Python легко розширюється за допомогою нових функцій і типів даних, що виконуються в C або C ++ (або інших мовах, які можна викликати з C). Python дозволяє писати програми компактними та зручними.

Програми, що написані на Python, зазвичай набагато коротші за еквівалентні програми на C, C ++ або Java, з кількох причин:

- високорівневі типи даних дозволяють виконати складні операції в одному виразі;
- групування коду здійснюється за допомогою відступу замість дужок;
- жодних оголошень змінних або аргументів не потрібно [19].

Django - це високорівневий Python Web фреймворк, який сприяє швидкій розробці та чистому, прагматичному дизайну. Фреймворк побудований досвідченими розробниками, тому він дбає про велику частину складностей веб-розробки, що дає змогу розробникам зосередити увагу на написанні додатка без необхідності винаходити колесо. Даний проект є безкоштовним та знаходиться у вільному доступі. Переваги фреймворка:

- висока швидкість розробки;
- досить високий рівень безпеки застосунків, за рахунок реалізованих у фреймворку засобів;
- проста масштабованість

Всі ці можливості django, зробили даний фреймворк найбільш популярним для розробки веб застосунків мовою python[20].

JavaScript - це «безпечна» мова програмування загального призначення. Дана мова не забезпечує низькорівневими засобами роботи з пам'яттю та процесором, оскільки з самого початку розроблялася для браузерів, в яких ці функції не потрібні.

Інші ж можливості залежать від оточення в якому запущений JS. JavaScript у браузері може виконати все, що стосується маніпуляцій з сторінкою, взаємодії з користувачем і, навіть з сервером:

- створювати HTML-теги, видаляти існуючі, змінювати стилі елементів, приховувати, показувати елементи, тощо;
- реагувати на дії користувача;
- виконувати спілкування з сервером без перезавантаження сторінки ("AJAX");
- і ще багато чого іншого [21].

jQuery - jQuery - це швидка, невелика та багатофункціональна бібліотека JavaScript. Що допомагає робити такі дії, як переміщення HTML-документів та маніпулювання, обробка подій, анімації та Ajax запити, набагато простіше. Це досягається за допомогою простого API, який працює в безлічі браузерів. З поєднанням універсальності та розширюваності, jQuery важливе місце у розробці мовою JavaScript. [22].

Це нова версія мови HTML, з новими елементами, атрибутами і поведінкою. HTML5 - найновіший стандарт HTML.

Властивості:

- HTML - це стандартна мова розмітки для створення веб-сторінок;
- означає мову розмітки гіпертексту;

- описує структуру веб-сторінок за допомогою розмітки;
 - елементи є будівельними блоками HTML-сторінок;
 - елементи HTML представлені тегами;
 - теги HTML містять елементи вмісту, такі як "заголовок", "параграф", "таблиця" тощо;
- браузер не показує теги HTML, але використовує їх для відтворення вмісту сторінки [23].

Bootstrap - це відкритий набір інструментів для розробки HTML, CSS та JS. Що допомагає швидко перетворювати свої ідеї в проект або створювати все додаток з самого початку за допомогою змінних і міксинів Sass, адаптивної системи сітки, великих попередньо встановлених компонентів та потужних плагінів, побудованих на jQuery. [24].

scikit-learn – бібліотека машинного навчання, на основі мови Python. Дана бібліотека характеризується наступними параметрами:

- надає прості та ефективні інструменти для добування та аналізу даних;
- доступна для всіх і для багаторазового використання в різних ситуаціях;
- побудована на основі numpy, scipy і matplotlib;
- є відкритим проектом, комерційне використання – лише під ліцензією bsd [25].

Elasticsearch - це розподілений пошуковий та аналітичний двигун, який побудований на основі концепції REST, здатний вирішити весь час зростаючу кількість можливих задач. Існують клієнти для найбільш популярних мов програмування Python, PHP, .Net. Дане рішення вирізняється наступними перевагами:

- велика швидкість роботи з даними;
- може запускатися будь де, починаючи з ноутбука і закінчуючи величезним сервером з великою кількістю інформації;
- має високий рівень еластичності та забезпечує надійну роботу всіх кластерів;

– зберігає будь які типи даних, це можуть бути цифрові, текстові, геодані, структуровані або не структуровані[26].

PostgreSQL - це потужна СКБД, що є об'єктно-реляційною базою даних з відкритим кодом, яка використовує та розширює мову SQL. Також поєднує у собі багато функцій, які дозволяють безпечно зберігати та масштабувати найскладніші дані. «Народження» PostgreSQL відбулося 1986 р. Як частина проекту POSTGRES в Каліфорнійському університеті в Берклі, з тих пір історія цього проект нараховує більш ніж 30 років активного розвитку.

PostgreSQL заслужив чудову репутацію, завдяки своїй перевірній архітектурі, надійності, цілісності даних, великого набору функцій, розширюваності та належності проекту до open source. PostgreSQL працює на всіх основних операційних система і має потужні додатки, такі як популярне розширення для геопросторової бази PostGIS. Саме тому PostgreSQL став улюбленим засобом серед реляційних баз даних для багатьох людей та організацій [27].

MongoDB - це база даних із відкритим вихідними кодом, на основі документів, що призначена для простоти розробки та масштабування. Ця база має наступні переваги:

а) забезпечує високу продуктивність збереження даних. зокрема:

- 1) підтримка вбудованих моделей даних зменшує активність вводу-виводу в системі бази даних.
- 2) індекси підтримують швидші запити і можуть включати ключі від вбудованих документів та масивів.

б) підтримує багату мову запитів для підтримки операцій читання та запису (crud), а також:

- 1) агрегація даних;
 - 2) текстовий пошук та геопросторові запити;
- в) механізм реплікації mongodb, який називається реплікатором, надає:
- 1) автоматичну відмово стійкість;
 - 2) резервування даних;

г) забезпечує горизонтальну масштабованість як частину своєї основної функціональності:

- 1) sharding, для поширення даних по кластерах машин;
- 2) підтримка створення зон даних на основі ключа shard[28].

3.2 Архітектура програмного забезпечення

3.2.1 Діаграма класів

Оскільки в фреймворку django використовується ORM (об'єктно реляційне відображення), то класи наведені в діаграмі класів майже в повній мірі відповідають таблицям в ER-діаграмі. Також в мові програмування python всі класи за замовчуванням унаслідуються від object, що зображено на діаграмі класів. Діаграма класів наведена у додатку А, плакат «Структурна схема класів».

3.2.2 Діаграма послідовностей

В таблиці 3.1 описані класи, які відображені в діаграмі послідовності та описано, за що відповідає кожний клас.

Таблиця 3.1 - Відповідальності класів.

Клас	Відповідальність
Користувач	Отримання інформації про матеріали, шляхом пошуку або отримання рекомендації. Та виконання певних дій з матеріалами: оцінка та коментування
Підсистема Веб	Клас, який виконує роль зв'язуючого елементу між користувачем, базами даних, та підсистемою виконання алгоритму. Містить в собі всю функціональну логіку, що виконується на веб сторінці.
Підсистема виконання алгоритму	Клас, який надає можливість обробляти інформацію про користувачів, об'єднувати їх у кластери та надавати їм рекомендації відповідно до кластерів

Продовження таблиці 3.1

Клас	Відповідальність
Сервер реляційної БД	Сервер БД, що надає доступ користувачу до інформації за допомогою БД. Оброблює запити, що надходять з підсистеми веб, які потребують управління даними.
Сервер NoSQL БД	Сервер БД, що зберігає модель отриману в підсистемі виконання алгоритму, та повертає її у випадку необхідності надати користувачеві рекомендацію.

На рисунку 3.1 вказано діаграму послідовності для отримання інформації в системі користувачем. На діаграмі показана взаємодія користувача підсистемою веб, підсистема веб в свою чергу взаємодіє з підсистемами сервера реляційної бази даних, алгоритму та NoSQL бази даних. Результатом взаємодії всіх цих систем, є підготовлені та відображені цільові пропозиції користувачу.

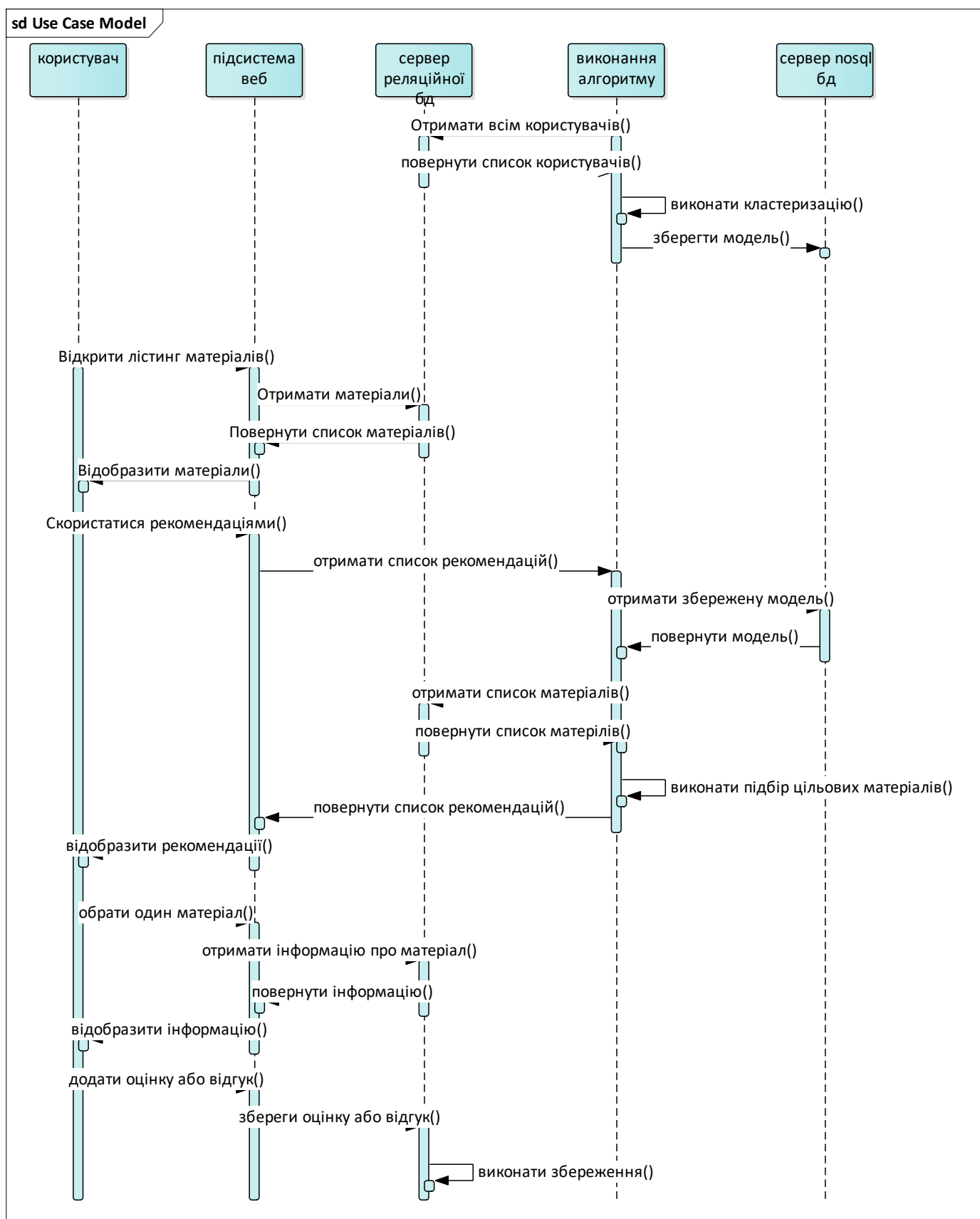


Рисунок 3.1 – Структурна схема послідовностей.

3.2.3 Архітектура Веб частини

Оскільки проект розробляється за допомогою django фреймворка то буде використовуватися архітектурний шаблон MVT, згідно якого система поділяється на три частини, що є взаємопов'язаними: модель даних, вигляд та шаблон. Даний шаблон наслідує шаблон MVC лише з певними змінами. Це засновується для зменшення зв'язності між частинами модулю, тобто для того, щоб частини залежали одна від одної мінімально.

Блок Модель (Model) — рівень доступу до даних. Даний рівень містить все про дані та описує: як отримати доступ до них, як перевірити їх, яка їх поведінка та які взаємозв'язки між даними.

Блок Відображення (View) — рівень бізнес-логіки. Цей рівень містить логіку, яка надає доступ до моделі та передає дані до відповідного По факту це просто рівень, який об'єднує модель та шаблон.

Блок Шаблон (Template) — рівень презентації даних з моделі, та даних, що були додані у блоці відображення. Цей рівень містить рішення, пов'язані з презентаціями: як щось відображати на веб-сторінці чи іншому документі. Зазвичай даний рівень реалізується за допомогою шаблону HTML.

На рисунку 3.2 вказано діаграму компонентів, що відображає дану архітектуру.

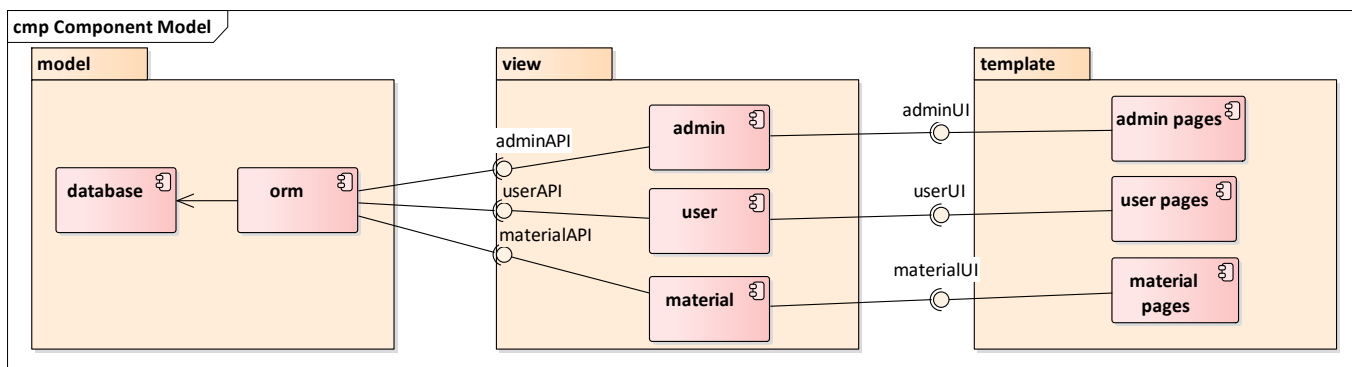


Рисунок 3.2 – Діаграма компонентів.

3.3 Інструкція адміністратора

3.3.1 Призначення підсистеми *back-office*.

Підсистема забезпечує користувачам з правами адміністратора можливість здійснювати управління всією системою. Адміністратор може управляти будь яким об'єктом в межах системи: матеріалом, коментарями, користувачем, групами користувача, тощо. Проте прямим обов'язком користувача з правами адміністратора є управління користувачами та групами користувачів.

3.3.2 Основні дії адміністратора

Для управління (створення, редагування, видалення) адміністратору варто перейти в один з наступних пунктів, що зображені на рисунку 3.3.

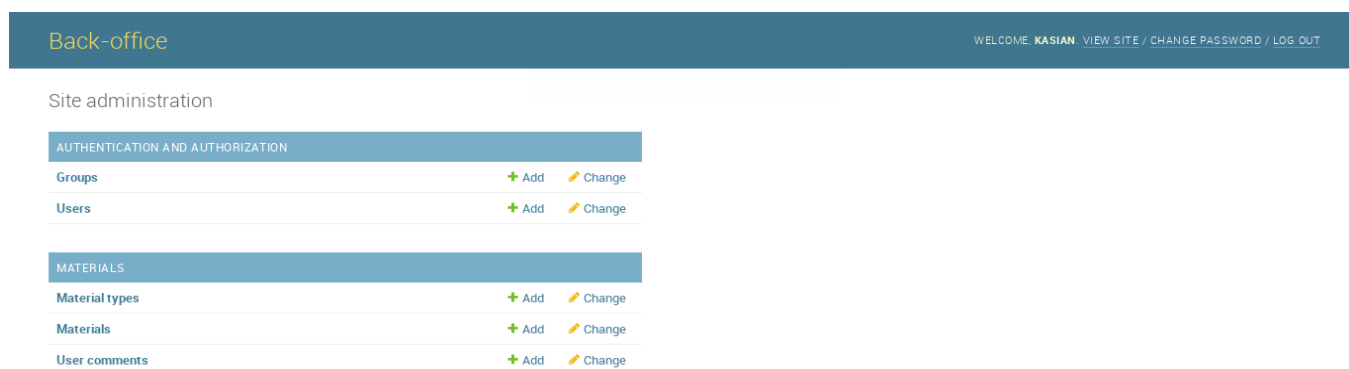


Рисунок 3.3 – Екранна форма списку сутностей відповідно до таблиць бази даних, якими може керувати адміністратор.

Після обрання певного об'єкту адміністратор потрапляє на сторінку лістингу об'єктів певної сутності. Лістинг зображено на рисунку 3.4.

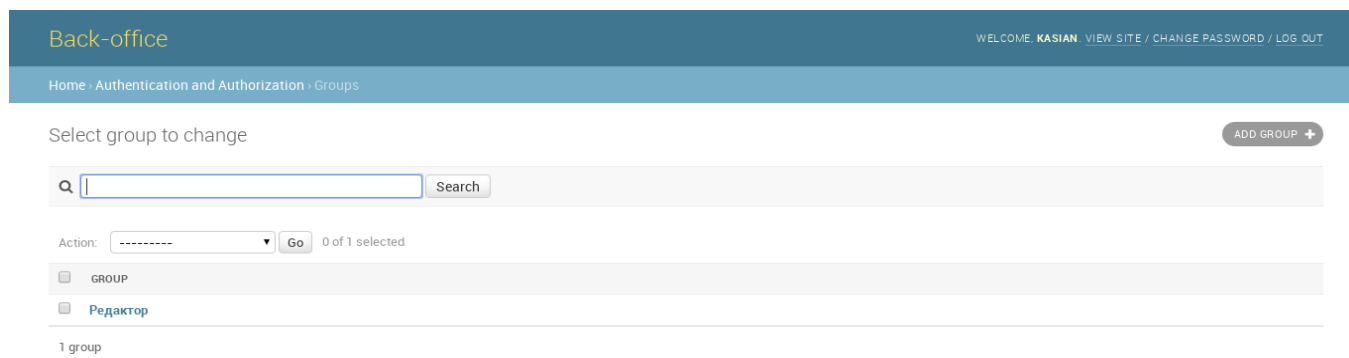


Рисунок 3.4 – Екранна форма списку об'єктів певної сутності.

Після обрання конкретного об'єкта, адміністратор має змогу вносити до нього зміни. Приклад, редагування зображено на рисунку 3.5.

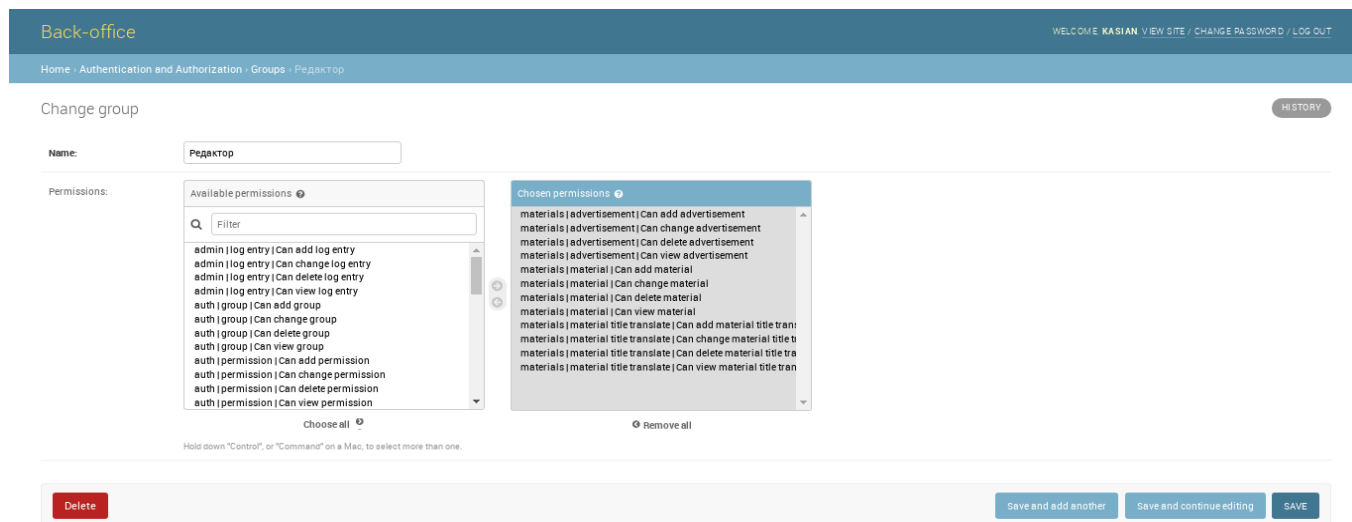


Рисунок 3.5 – Екранна форма редагування певного об'єкта на прикладі груп користувачів.

3.4 Інструкція редактора

3.4.1 Призначення підсистеми *back-office*

Роль редактора та підсистема розроблені для наповнення системи необхідними матеріалами. Користувач з правами редактора має змогу, додавати, редагувати та видаляти матеріали. Також даний користувач може додавати рекламу до матеріалу або ж відгук від імені редакції.

3.4.2 Основні дії редактора

Послідовність дій редактора така ж сама, відрізняється лише список сутностей які може він змінювати. На рисунку 3.6 зображено список сутностей, які може змінювати редактор.



Рисунок 3.6 – Екранна форма списку сутностей відповідно до таблиць бази даних, якими може керувати редактор.

Після обрання конкретного матеріалу, або при створенні нового відкривається форма редагування певного об'єкту. На рисунку 3.7 зображено створення нового матеріалу.

The screenshot shows a web interface for adding a new material. At the top, there is a dark blue header with the text 'Back-office' on the left and 'WELCOME ВОЛОДИМИР / VIEW SITE / CHANGE PASSWORD / LOG OUT' on the right. Below the header is a light blue breadcrumb trail: 'Home > Materials > Materials > Add material'. The main content area is titled 'Add material'. It contains several input fields: 'Image:' with a 'Choose File' button and 'star.jpg' text; 'Title original:' with a text box containing 'Збірка віршів'; 'Description:' with a large text area; and 'Technical description:' with a smaller text area. A green circular icon with a 'G' is visible in the bottom right corner of the description area.

Рисунок 3.7 – Екранна форма створення нового матеріалу.

На рисунку 3.8 зображено можливість додавання реклами та перекладу назви для матеріалу.

The screenshot shows two forms. The first form, titled 'MATERIAL TITLE TRANSLATES', has a sub-header 'Material title translate: #1'. It includes fields for 'Language:' (set to 'en') and 'Translate:' (set to 'Collection of poems'). Below these is a link '+ Add another Material title translate'. The second form, titled 'ADVERTISEMENTS', has a sub-header 'Advertisement: #1'. It includes fields for 'Picture:' (with a 'Choose File' button and 'No file chosen' text), 'Title:', 'Text:', 'Reference:', and 'Author:' (a dropdown menu). It also has 'Date start:' and 'Date end:' fields, each with a 'Today' button and a calendar icon. Below these are notes: 'Note: You are 2 hours ahead of server time.' and a link '+ Add another Advertisement'.

Рисунок 3.8 – Екранна форма додавання реклами та перекладу до матеріалу.

3.5 Інструкція користувача

3.5.1 Призначення підсистеми front-office.

Дана підсистема є основною, адже саме через цю підсистему відбувається взаємодія з користувачем. Дій користувача в 1-й версії відносно обмежені. Користувач може переглядати типи доданих матеріалів, списки матеріалів відповідно до типу, переглядати рекомендації, переглядати деталі матеріалу та оцінювати його.

3.5.2 Основні дії користувача

При потраплянні на сайт авторизований користувач може бачить список доданих типів матеріалів, та може обрати один з них, це зображено на рисунку 3.9.

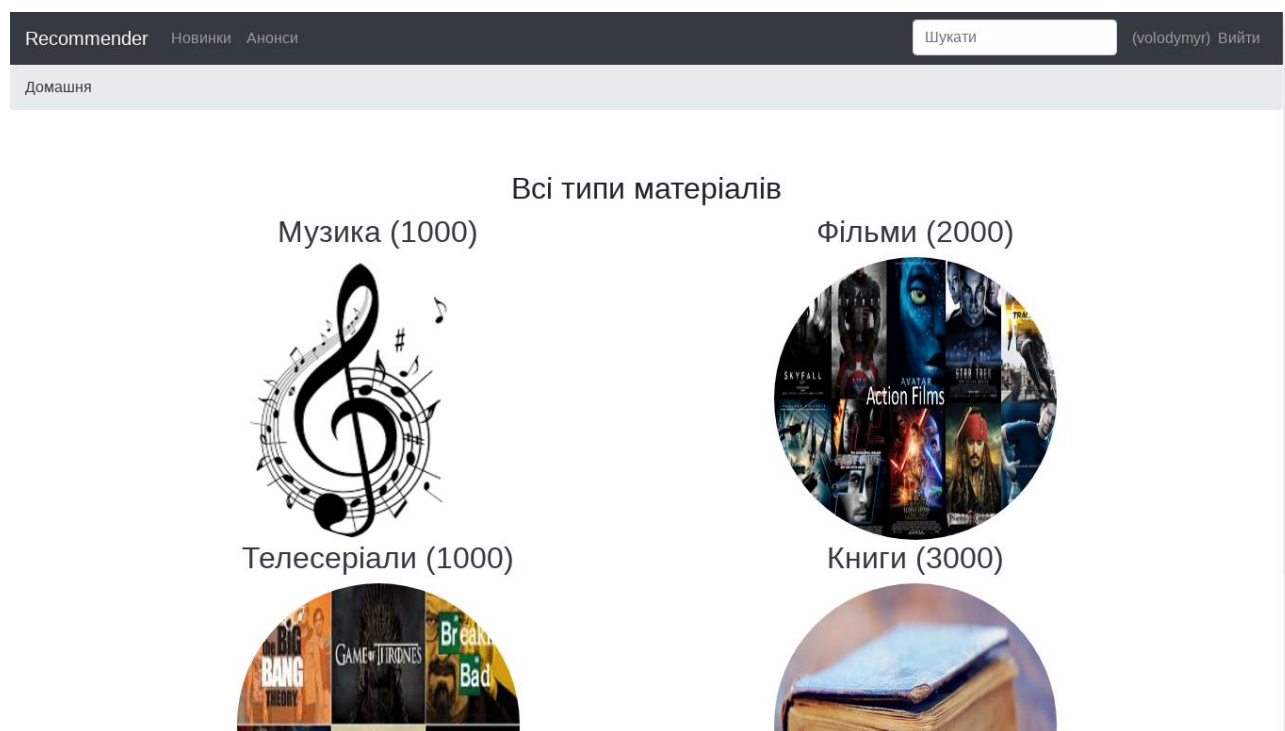


Рисунок 3.9 – Екранна форма перегляд типів матеріалів.

Після обрання типу матеріалів, користувач потрапляє на сторінку зі списком матеріалів певного типу з коротким описом. Після чого користувач може перейти до деталей матеріалу або переглянути рекомендації певного типу. На рисунку 3.10 зображено список фільмів, а на рисунку 3.11 список книг, для інших типів не наводиться до вони мають аналогічний вигляд.

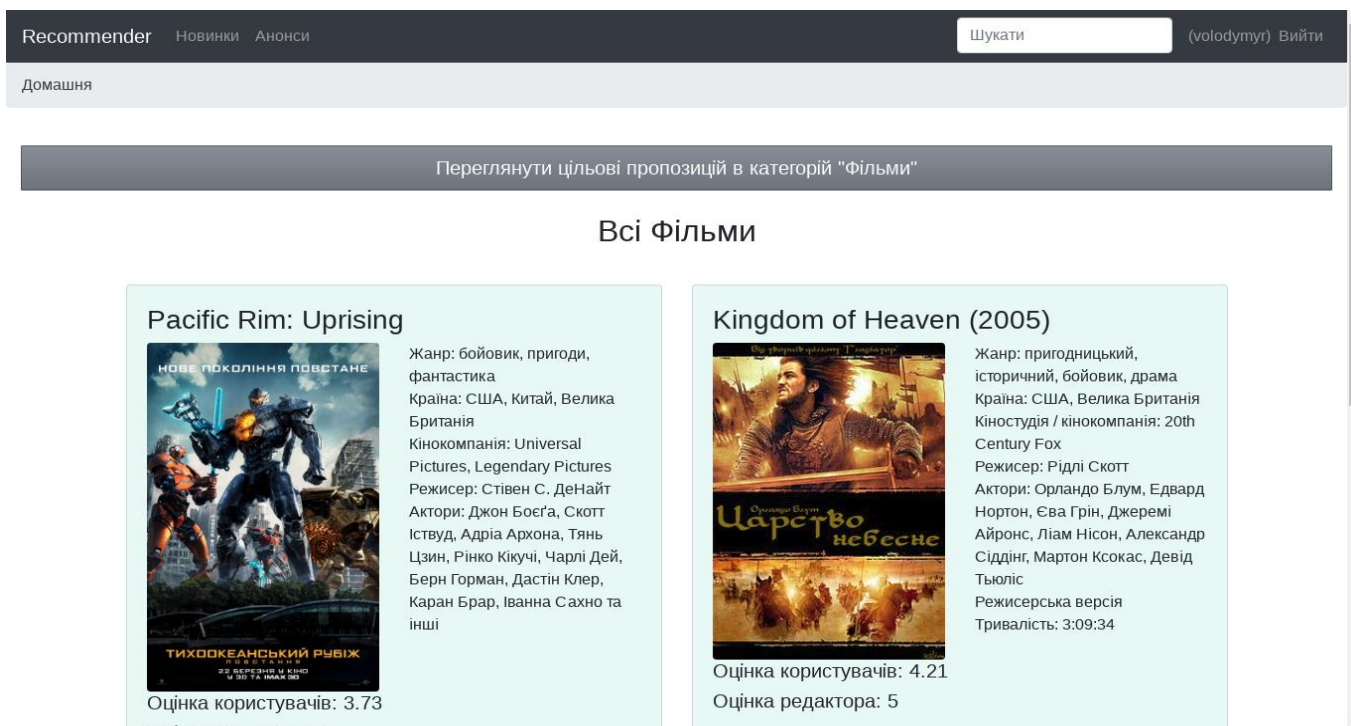


Рисунок 3.10 – Екранна форма списку фільмів.

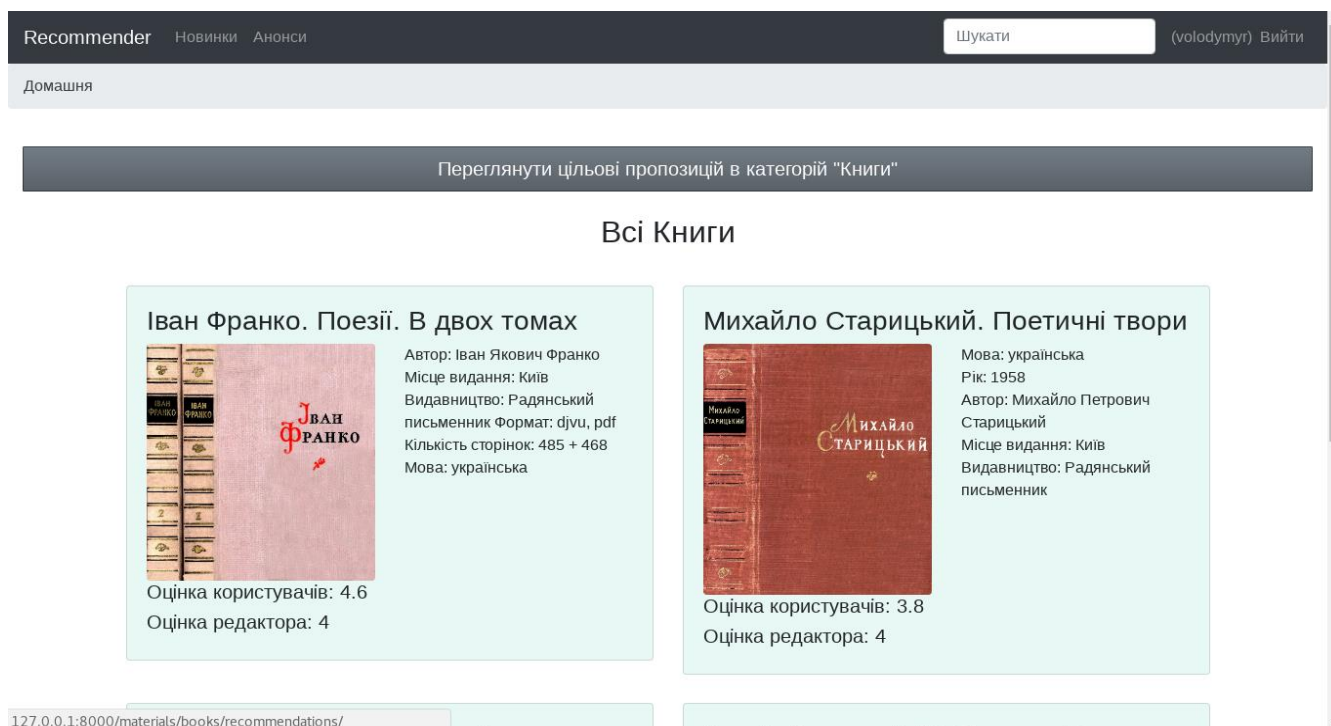


Рисунок 3.11 – Екранна форма сторінки списку книг.

При обранні рекомендацій, користувач так само бачить список матеріалів певного типу, які були підібрані системою, як такі що можуть сподобатися користувачу. На рисунку 3.12 зображено список цільових пропозицій.

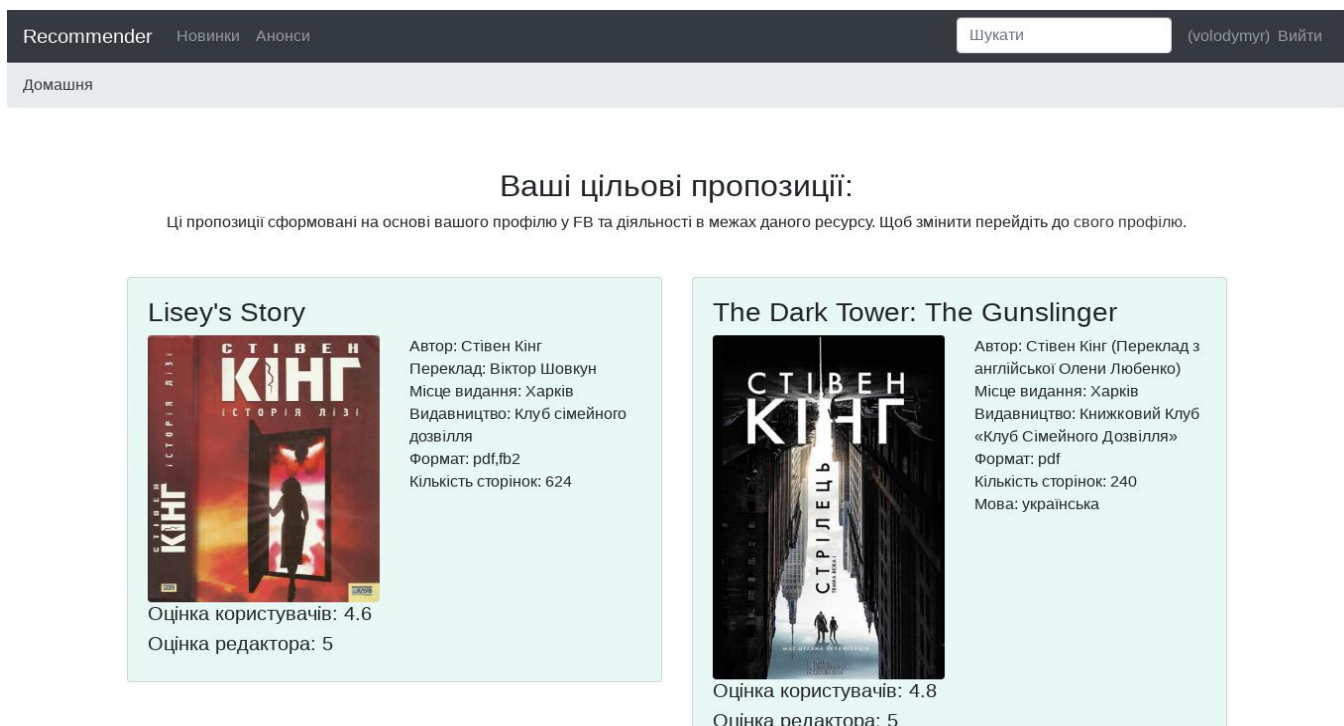


Рисунок 3.12 – Екранна форма сторінки списку цільових пропозицій.

При обранні конкретного матеріалу користувач потрапляє на сторінку з детальним описом з можливістю не лише переглянути, а поставити власну оцінку. Дані можливості зображено на рисунку 3.13.



Рисунок 3.13 – Екранна форма сторінки деталей матеріалу.

3.6 Опис технічного забезпечення

Вимоги до технічного забезпечення впливають зі складності поставлених задач.

Оскільки серверна частина складається з достатньої кількості програмних продуктів, та повинна обробляти досить велику кількість даних, наведемо вимоги до серверу:

- процесор тактова частота, якого дорівнює 2,5 ГГц;
- оперативна пам'ять повинна бути 8 ГБ;
- об'єм HDD повинен дорівнювати 200Гб і більше;
- повинна бути можливість налаштовувати сервер віддалено.

На стороні клієнта виконується порівняно небагато дій, але сучасні фреймворки та бібліотеки все ж можуть бути доволі затратними, тому сформуємо вимоги до клієнтської машини:

- процесор, тактова частота якого ≥ 1 ГГц;
- об'єм оперативної пам'яті ≥ 2 Гб;
- встановлений будь який сучасний браузер;
- доступ до мережі Інтернет.

Також для роботи найбільш зручного використання сервісу необхідні периферійні пристрої, у випадку використання стаціонарного комп'ютера:

- монітор;
- мишка;
- клавіатура.

Наведено діаграму розгортання у додатку А, плакат «Структурна схема діаграми розгортання». На діаграмі зображено взаємодію підсистем серверної та клієнтської частини інформаційної системи.

Висновок до розділу

В розділі обрано та описано засоби розробки. При обранні засобів відштовхувався від вимог до розробки та її мети.

Також, наведено архітектуру програмного продукту, зокрема діаграму класів, діаграму компонентів та діаграму розгортання. Також наведено опис технічного забезпечення необхідного для коректної роботи сервера та клієнтської машини.

Наведено інструкцій для кожної з ролей: адміністратора, редактора, користувача.

Даний розділ покликаний допомогти розробнику та кінцевому користувачу у розгортанні та використанні відповідно, готового програмного продукту.

4 РОЗРОБКА СТАРТАП-ПРОЕКТУ

4.1 Опис ідеї проекту (товару, послуги, технології)

В таблиці 4.1 наведена основна ідея проекту, а також основний та додаткові напрямки застосування розробки.

Таблиця 4.1. Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Інформаційна система яка є універсальною базою даних мультимедійних розваг (фільми, музика, телепередачі), друкованих матеріалів, тощо. До системи включена підсистема підготовки цільових пропозицій кожному користувачу, яка враховує вподобання користувача, а також підсистема, яка групує користувачів, які мають схожу поведінку в соціальних мережах, тому (можуть бути односторонніми), а також визначає ступінь приналежності користувача для кожної з таких груп.	1. Система підготовки цільових пропозицій	Поліпшення досвіду користування ресурсом, полегшення підбору необхідних матеріалів, відкриття чого нового
	2. Система підбору реклами відповідно до цільових пропозицій	Отримання реклами, яка зв'язана лише з тим що може бути цікавим користувачу
	3. Підсистема збору відгуків та оцінки матеріалів від аудиторії різного віку.	Частковий (непрямий) вплив на матеріал який створюється, за рахунок того що активний діяльність щодо оцінки матеріалів, допоможе врахувати помилки в подальшій роботі.
	4. Підсистема аналізу потреб та бажань користувачів	Дуже схоже на попереднє, за виключенням того, що така підсистема допоможе обрати напрям авторам матеріалу ще до початку роботи.

Система має наступних конкурентів:

- Netflix — сервіс онлайн прокату фільмів, з підсистемою надання рекомендацій мультимедійних матеріалів;
- IMDb — база даних напевне всіх фільмів, з системою, рекомендації користувачам фільмів ;

– Youtube — сервіс для перегляду відео контенту, є напевне найбільш популярним сервісом в даній сфері, та має підсистему, яка надає користувачеві рекомендації на основі його діяльності в даному сервісі.

Таблиця 4.2. Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№ п/п	Техніко-економічні характеристики ідеї	(потенційні) товари/концепції конкурентів				W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
		Мій проект	Netflix	IMDB	Youtube			
1	Вартість обслуговування	На початковому етапі - невелика	Відсутня інформація. Припущення - висока	Відсутня інформація. Припущення – дуже висока	Відсутня інформація. Припущення – дуже висока		+	
2	Вартість експлуатації	Близько 100\$ в місяць на початковому етапі	Відсутня інформація.	Відсутня інформація.	Відсутня інформація.		+	
3	Безвідмовність	Декілька екземплярів та баласувальник навантаження забезпечує безвідмовну роботу	Дуже висока	Досить висока	Дуже висока		+	

Продовження таблиці 4.2

№ п/п	Техніко- економічні характерис- тики ідеї	(потенційні) товари/концепції конкурентів				W (слаб- ка сторо- на)	N (нейт- ральн- а сторо- на)	S (сильна сторона)
		Мій проект	Netflix	IMDB	Youtube			
4	Оплата праці	На початковому етап система є стартапом. Тому її будуть розвивати зацікавлені люди	Дуже висока	Відсутн- я інформ- ація. Припу- щення – висока	Дуже висока			+
5	Зручність користуван- ня	На початковому етапі не має потужного інфтерфейсу, але відсутня розробка дизайну з врахуванням UX	Висока	Середн- я	Дуже висока	+		
6	Дизайн ресурсу	На початковому етапі не розроблявся	Висока	Висока	Середн- я	+		
7	Відповідніс- ть патернами дизайну	Середня	Висока	Висока	Середн- я	+		
8	Відповідніс- ть тенденціям дизайну	Середня	Висока	Середн- я	Дуже висока	+		

Продовження таблиці 4.2

№ п/п	Техніко- економічні характерис- тики ідеї	(потенційні) товари/концепції конкурентів				W (слаб- ка сторо- на)	N (нейт- ральн- а сторо- на)	S (сильна сторона)
		Мій проект	Netflix	IMDB	Youtube			
9	Безпека даних користувач- а	Висока (забезпечує- ться сучасними засобами розробки, та дотримання- м найкращих практик при розробці бек-енду)	Дуже висока	Дуже висока	Середн- я		+	
10	Відмово- стійкість системи	За рахунок розподілу системи на мікросервіси , дуже мала ймовірність що вийде вся система з ладу	Дуже висока	Дуже висока	Дуже висока		+	
11	Інтерфейс, що «змушує» користувач- а залишатися на сайті	На головній сторінці виводяться рекомендації , які користувачу можуть бути цікавими	Розробл- ені алгорит- ми, які підбира- ють навіть кольоро- ву палітру під користу- вача	Відсутн- і елемент- и які могли б приверн- ути увагу користу- вача	Виводя- ться популяр- ні відео, відео, що можуть сподоба- тися користу- вачу тощо		+	

Продовження таблиці 4.2

№ п/п	Техніко- економічні характерис- тики ідеї	(потенційні) товари/концепції конкурентів				W (слаб- ка сторо- на)	N (нейт- ральн- а сторо- на)	S (сильна сторона)
		Мій проект	Netflix	IMDB	Youtube			
12	Широка предметна область	Використанн я для мультимедій них та друкованих матеріалів	Викори- стання для мульти- медійни- х матеріа- лів	Викори- стання для мульти- медійни- х матеріа- лів	Викори- стання для мульти- медійни- х матеріа- лів			+
13	Врахування особливост ей користувач а	Система в більшій мірі націлена на користувачів з України	Систем а на дуже придатн а для викорис- тання більшо- ю частино ю жителів України	Систем а містить перекла- ди лише в деяких місцях. Тобто система націлен а на користу- вачів з англомо- вних країн	Систем а зручна для всіх користу- вачів.		+	

4.2 Технологічний аудит ідеї проекту

Розробка технологій чи системи, потребує аналізу наявних засобів та проблем з якими можна зіткнутися, у випадку відсутності засобів. Тому, в таблиці 4.3 наведено технологічний аудит ідеї.

Таблиця 4. 3. Технологічна здійсненність ідеї проекту

№ п/п	Ідея проекту	Технології реалізації	Наявність технологій	Доступність технологій
1	Отримання даних про діяльність користувача в соціальній мережі facebook	Відкритий прикладний програмний інтерфейс facebook, який водночас є досить обмеженим	Наявна	Доступна в достатньому обсязі
2	Об'єднання користувачів у групи	Методи та засоби кластеризації	Необхідно розробити	Засоби є доступними
3	Підготовка цільових пропозицій	Методи та засоби підготовки цільових пропозицій	Необхідно розробити	Засоби є доступними
4	Програмний продукт у вигляді інформаційної системи	Програмні засоби для розробки системи та описані вище методи	Необхідно розробити	Використовуються засоби, що є вільними для використання
Обрані технології реалізації ідеї проекту: python, javascript, html, django, jquery, scikit, requests, postgresql, mongodb, elasticsearch				

4.3 Аналіз ринкових можливостей запуску стартап-проекту

Стартап проект розробляється, для того щоб бути впровадженим. Саме тому необхідно визначити ринкові можливості проекту під час впровадження. У таблиці 4.4 наводиться аналіз попиту, обсяг та динаміку розвитку ринку.

Таблиця 4.4. Попередня характеристика потенційного ринку стартап-проекту

№ п/п	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, од	5
2	Загальний обсяг продаж, грн/ум.од	-
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу (вказати характер обмежень)	1. Закони про збір, обробку та зберігання особистих даних
5	Специфічні вимоги до стандартизації та сертифікації	Вимоги відсутні
6	Середня норма рентабельності в галузі, %	~40-60%

У таблиці 4.5 наводяться визначені потенційні клієнти, а також їх характеристика.

Таблиця 4.5. Характеристика потенційних клієнтів стартап-проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1	Необхідність спрощення підбору ресурсів звичайному користувачу	Активні користувачі всесвітньої мережі Інтернет	Поведінка залежить в більшості від професійної діяльності, хобі та віку	1. Збереження конфіденційності приватних даних. 2. Отримання пропозицій, що будуть відповідати очікуванням
2.	Аналіз поведінки користувача для покращення рекламних пропозицій	Рекламодавці	Поведінка залежить від продукту, що рекламується	1.Отримання інформації про групи користувачі, чим схожі та чим відрізняються, про їх вподобання, та особливості поведінки в системі

Результатами аналізу ринкового середовища є описані фактори загроз та можливостей. У таблиці 4.6 наведені фактори, що перешкоджають ринковому впровадженню.

Таблиця 4.6. Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1	Конкуренція з технологічними гігантами	Проблеми в просуванні системи	Запуск системи Україні з подальшим розширенням кордонів
2	Відсутність зацікавленості у потенційних користувачів	Система може не викликати довіри у нових користувачів	Підвищення якості продукту, поступовий аналіз ринку та проведення маркетингових дій для підвищення популярності
3	Зміни законів, щодо зберігання та обробки персональних даних.	Може спричинити видалення необхідних даних.	Розробка альтернативних методів. Адаптація системи до чинного законодавства

У таблиці 4.7 наведені фактори, що сприяють ринковому впровадженню.

Таблиця 4.7. Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Отримання можливості легально поширювати матеріали	Отримання дозволів реалізації в системі перегляду відеоматеріалів, прослуховування аудіо, перегляд друкованих матеріалів.	Реалізація даного функціоналу.
2	Здобуття популярності на ринках за межами України	Користувачі судніх держав також можуть зацікавитися в системі.	Адаптація системи під ринки з найбільшою кількістю користувачів.
3	Зацікавленість інвесторів	Можливість отримати додаткові інвестиції.	Розширення штату розробників та покращення функціональних можливостей системи

У таблиці 4.8 наведено загальні риси конкуренції на ринку.

Таблиця 4.8. Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Вказати тип конкуренції - монополія	Монополія технологічних гігантів, які є чітко виділеними лідерами	Запуск системи на українському ринку з поступовим розширенням та реалізація додаткового функціоналу, якого не мають конкуренти
2. За рівнем конкурентної боротьби - міжнародний	Монополісти є міжнародними компаніями	Вихід на міжнародний ринок, після закріплення на локальному
3. За галузевою ознакою - внутрішньогалузева	Всі продукти є програмними і майже всі розміщені в мережі інтернет	Пошук нових ідей та втілення їх
4. За характером конкурентних переваг - нецінова	Нецінова, бо звичайному користувачу не потрібно платити за використання системи	Орієнтація на потреби користувачів

У таблиці 4.9 наведено більш детальні риси конкуренції на ринку.

Таблиця 4.9. Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари замітники
	IMDB	1.Google play music 2.Apple music 3.Youtube 4. Rozetka	Постачальниками є кінокомпанії, видавництва, тощо	Будь – які користувачі мережі інтернет	Кожен з продуктів конкуренті в є частковим заміником
Висновки :	На українському ринку, і в сферах відмінних від кіно, прямих конкурентів немає	Можливість виходу на ринок досить висока, потенційні конкуренти вже мають продукти зі схожими функціями, про розширення поки не чути.	Постачальники не диктують умови ринку, адже достатньо придбати права на публікацію матеріалу,лише у випадку публікації повного матеріалу.	Клієнти диктують вимоги до ринку. Адже саме на них спрямовані всі сервіси	Поки всі продукти спрямовані на якусь конкретну область, що дає змогу реалізувати і випустити власний продукт.

На основі вище наведених таблиці сформульовано фактори конкурентоспроможності, та наведено у таблиці 4.10.

Таблиця 4.10. Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування
1	Спрямованість системи на декілька предметних областей	Продукти спрямовані на надання цільових пропозицій лише в певній області (фільми або музика, тощо). Розроблена система на початковому етапі має 4 області: фільми, музика, книги, телесеріали
2	Рекомендації користувачу в кожній сфері	Завдяки тому, що в системі використовується кластеризація ми можемо підготувати користувачу пропозицій, навіть в тій області, про яку у нас немає відомостей.
3	Рекомендації, які базуються на вподобаннях однодумців	Система враховує діяльність користувача в межах ресурсу, а також діяльність в соціальних мережах, що дає змогу «віднести» його до групи однодумців.

Продовження таблиці 4.10

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
4	Можливість надання цільової реклами	Рекламодавці можуть врахувати інформацію про кластери користувачі, і підготувати рекламу відповідно до цього опису і показати її лише для «необхідних» користувачів.
5	Закладена можливість розширення системи	В системі передбачено додання нових модулів та функціоналу.
6	Адаптація системи під певний локальний ринок	Система передбачає локалізацію.

На основі факторів конкурентоспроможності порівнюємо сильні та слабкі сторони системи. Порівняння наведено у таблиці 4.11.

Таблиця 4.11. Порівняльний аналіз сильних та слабких сторін «інформаційної технології кластеризації користувачів для підготовки цільових пропозицій»

№ п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з системою рекомендацій на основі поведінкової моделі користувача						
			-3	-2	-1	0	+1	+2	+3
1	Спрямованість системи на декілька предметних областей	20		+					
2	Рекомендації користувачу в кожній сфері	15			+				
3	Рекомендації, які базуються на вподобаннях однодумців	10				+			
4	Можливість надання цільової реклами	10					+		
5	Закладена можливість розширення системи	10					+		
6	Адаптація системи під певний локальний ринок	10			+				

На основі вище наведених таблиць підведемо підсумки ринкового аналізу, та сформуємо їх у вигляді SWOT- аналізу (таблиця 4.12).

Таблиця 4.12. SWOT- аналіз стартап-проекту

Сильні сторони:	Слабкі сторони:
Спрямованість системи на декілька предметних областей Рекомендації користувачу в кожній сфері Адаптація системи під певний локальний ринок	Погано розроблений дизайн Низька зручність системи «Слабка» реалізація засобів цільової реклами
Можливості:	Загрози:
Отримання можливості легально поширювати матеріали Здобуття популярності на ринках за межами України Зацікавленість інвесторів	Конкуренція з технологічними гігантами Відсутність зацікавленості у потенційних користувачів Зміни законів, щодо зберігання та обробки персональних даних.

На основі SWOT- аналізу розроблено альтернативи ринкової поведінки та орієнтовний час виведення проекту на ринок (таблиця 4.13).

Таблиця 4.13. Альтернативи ринкового впровадження стартап-проекту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Запуск системи лише з основним функціоналом	Висока ймовірність. Інвестори зможуть оцінити ідею на практиці	3 місяці
2	Поліпшення основних можливостей системи на основі реакцій користувачів	Відсутня	2 місяць
3	Реалізація іншого задуманого функціоналу системи в тому числі для рекламодавців	Висока ймовірність знову отримати інвестиції, а також отримувати кошти від рекламодавців	8 місяців

Обрано 1 альтернативу, оскільки вона має низький термін реалізації та високу ймовірність отримати ресурси.

4.4 Розроблення ринкової стратегії проекту

Для ринкової стратегії потрібно виконати батокроковий аналіз. Спочатку визначимо стратегії охоплення ринку (таблиця 4.14).

Таблиця 4.14. Вибір цільових груп потенційних споживачів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Люди молодого віку	Середня	Невисокий	На момент запуску низька	Висока складність залучення
2	Люди середнього віку	Середня	Середній	На момент запуску низька	Висока складність залучення
3	Люди літнього віку	Низька	Низький	На момент запуску низька	Висока складність залучення
4	Люди будь якого віку, які є користувачами глобальної мережі інтернет	Висока	Вище середнього	На момент запуску низька	Середня складність залучення
5	Люди які заробляють в інтернеті (блогери, продавці, ...)	Середня	Низький	Середня	Середня складність залучення. Але високі ризики з контентом
Які цільові групи обрано: на початковому етапі обрано людей будь якого віку, які є користувачами мережі інтернет, пізніше буде залучено людей, які заробляють в інтернеті					

Оскільки, доведеться працювати з декількома сегментами то, використаємо стратегію диференційованого маркетингу та наведемо базову стратегію розвитку.

Таблиця 4.15. Визначення базової стратегії розвитку

№ п/п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
	Запуск якомога швидше, застосунок з основним функціоналом	Диференційований маркетинг	Вища швидкість надання пропозицій, а також надання пропозицій в більшій кількості сфер	Стратегія диференціації

У таблиці 4.16 описано стратегію конкурентної поведінки.

Таблиця 4.16. Визначення базової стратегії конкурентної поведінки

№ п/п	Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки
1	Частково	Забирати існуючих на початковому етапі не буде. Шукати нових споживачів буде	Компанія буде досліджувати найкращі практики в розробці веб-сервісів. Конкуренти можуть інколи, задавати ці найкращі практики.	Стратегія лідера на локальному ринку і з часом стратегія виклику лідера на міжнародному

На основі описаних вище стратегій опишемо стратегію позиціонування продукту.

Таблиця 4.17. Визначення стратегії позиціонування

№ п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап проекту	Вибір асоціацій, які мають сформулювати комплексну позицію власного проекту (три ключових)
1	Зручний та швидкий у використанні	Стратегія диференціації	Сучасні методи машинного навчання	Швидкий та якісний підбір пропозицій
2	Надійно захищає особисті дані користувача	Стратегія диференціації	Сучасні засоби захисту даних	Надійний захист особистої інформації
3	Надає релевантні пропозиції і правдиву інформацію	Стратегія диференціації	Розробка та постійне удосконалення методу підготовки цільових пропозицій та модерація інформації	Постійне поліпшення методу підготовки пропозицій

4.5 Розроблення маркетингової програми стартап-проекту

Для того щоб сформувавши маркетингову концепцію продукту, підсумуємо результати аналізу конкурентоспроможності продукту (таблиця 4.18).

Таблиця 4.18. Визначення ключових переваг концепції потенційного товару

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Висока релевантність пропозицій	Алгоритм надання пропозицій весь час покращується	Надання релевантних об'єктів в різних областях
2	Актуальність інформації	Робота редакторів забезпечує актуальну інформацію про матеріали	Постійне оновлення матеріалів, відповідальними за це людьми
3	Врахування зміни вподобань	Кластеризація відбувається щоночі, що дозволяє відреагувати на зміну вподобань	Врахування будь яких змін в «поведінці» користувача

У таблиці 4.19 описані ідея продукту, його складові та особливості процесу надання.

Таблиця 4.19. Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Програмний продукт у вигляді веб застосунку, який використовується для надання користувачам цільових пропозицій, щодо матеріалів (фільми, книги, ...) на основі його діяльності в соціальній мережі facebook та в межах даного застосунку		
II. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	1. Висока релевантність наданих пропозицій	М	Тх/Тл
	2. Актуальність інформації		
	3. Захищеність особистих даних користувача		
	Система була повністю протестована та відповідає зазначеним параметрам якості. Наявна документація		
Веб сайт, з користувацьким та адміністраторським інтерфейсом			
Марка: MultiRecommender			
III. Товар із підкріпленням	Програмний продукт		
	Програмний продукт		
За рахунок чого потенційний товар буде захищено від копіювання: захищений програмний код, а також використання ліцензії			

Вартість на товар визначається вартістю реклами, адже саме реклама є джерелом прибутку в даній системі. Тому вкажемо в таблиці 4.20 ціни відповідно до реклами.

Таблиця 4.20. Визначення меж встановлення ціни

№ п/п	Рівень цін на товари замітники, грн/перегляд	Рівень цін на товари аналоги, грн/перегляд	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
1	1	3-4	Середній - високий	Оскільки в системі високий відсоток унікальних користувачів, а також реклама є цільовою то вартість її буде в межах 3-5 грн/перегляд, в залежності від розмірів кластерів яким вона пропонується

У таблиці 4.21 наведена оптимальна система збуту.

Таблиця 4.21. Формування системи збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
1	Придбання консультацій з редактором, який допоможе з визначенням кластерів, яким надаватиметься реклама. А також оплата власне реклами	-	До користувачів, яким необхідна реклама	Власний веб-сайт

У таблиці 4.22 наведено розроблену концепцію маркетингових комунікацій.

Таблиця 22. Концепція маркетингових комунікацій

№ п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
1	Знають тематику своєї реклами, можуть бути не впевненими щодо аудиторії	Веб-сайти, месенджери, соціальні мережі	Релевантні пропозицій користувачам, відповідно релевантна реклама. Захищеність даних користувачів, відповідно зменшення можливостей у конкурентів на антирекламу	Показати можливість користувача отримувати актуальні та релевантні пропозиції, а рекламодавцям надавати цільові рекламні пропозиції	Система для отримання актуальних та релевантних пропозицій

Висновок до розділу

У розділі визначена доцільність розробки програмного продукту та описано стратегію виведення його на ринок.

Даний продукт є універсальною базою даних мультимедійних розваг (фільми, музика, телепередачі), друкованих матеріалів, тощо, який немає прямих конкурентів, проте існуючі конкуренти є досить небезпечними.

Тому, стартап вирішено запускати в першу чергу в Україні. Адже ринок в Україні є досить привабливим для даної системи.

Визначено, що проблем з реалізацією не має виникати. Також визначено слабкі та сильні сторони проекту, де слабкі сторони зникають з поступовою розробкою системи.

Обрано альтернативу, яка передбачає запуск системи лише з основним функціоналом, для отримання інвестицій.

Обрано цільові групи, які складаються з людей будь якого віку, що користуються Інтернетом, а також з людей, що заробляють за допомогою інтернету.

Відповідно до цього обрано диференційований маркетинг, як стратегію охоплення ринку та стратегію диференціації, як стратегію розвитку.

Розроблено маркетингову стратегію, яка полягає у продажі рекламних оголошень.

Отже, спираючись на даний аналіз можна передбачити, що проект успішно вийде на локальний ринок, після чого можна буде будувати плани виходу на міжнародний.

ВИСНОВКИ

1) Реалізовано підсистеми збору даних в межах розробленого продукту, та з соціальної мережі facebook.

2) Проведено аналіз вимог до системи, та обрано метод колаборативної фільтрації для підготовки цільових пропозицій. Даний метод модифіковано шляхом, використання в замість функцій схожості нечіткої кластеризації.

3) Оглянуто методи та обрано метод, для якого не потрібні зарання відомі центри кластерів, та навіть кількість кластерів. Обрано метод поширенні близькості, хоча сам по собі він не є нечітким. Тому даний метод було модифіковано

4) Модифікація методу полягає у використанні матриці схожості замість матриці відстаней, а також у подальшому збереженні схожості користувачів при переобчисленні центрів кластерів. Оцінка даної модифікації є майже еквівалентною, до інших методів кластеризації. З мінусів трохи довше виконання методу, яке компенсується сучасними засобами розробки. І тому зважаючи на метрики, а також на можливість виконання методу без заданих завчасно кластерів, цей метод є найкращим серед розглянутих для використання його в системі колаборативної фільтрації. Також ми можемо розглядати кластери, як узагальненого користувача, а міру належності до кластера, як міру схожості між поточним та узагальненим користувачем.

5) Даний метод було використано при розробці інформаційної системи з підсистемою підготовки цільових рішень відповідно до певної діяльності користувача в соціальній мережі facebook, а також його вподобань. Також модифікований метод кластеризації може використаний і в інших цілях, до прикладу визначення можливості прихильності користувача до певної політичної партії, тощо.

6) Для подальшого покращення інформаційної технології підготовки цільових пропозицій, потрібні певні статистичні, які покликана надати розроблена інформаційна система.

7) Реалізована інформаційна система виконує покладені на неї функції збору та аналізу даних для полегшення процесу вибору користувачем. Також, згідно аналізу дана, система може бути стартап проектом, там має шанс вийти на локальний ринок.

8) До подальших покращень системи можна віднести:

- а) оптимізація алгоритму, на основі отриманих статистичних даних;
- б) оптимізація швидкодії алгоритму кластеризації;
- в) розширення кількості параметрів отриманих з соціальної мережі facebook, наприклад шляхом отримання та аналізу текстів публікацій;
- г) розширення кількості параметрів, що збираються в системі;
- д) покращення інформаційної системи

9) За матеріалами дисертації було опубліковано 2 наукові роботи: 2 тез доповіді на конференціях [29-30].

ПЕРЕЛІК ПОСИЛАНЬ

1. Koutheair M. K. Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval / M. K. Koutheair, M. Jemni, O. Nasraoui. // Educational Technology & Society. – №12. – August 2008 . – pp. 241 – 245.
2. Lu J. A Personalized e-Learning Material Recommender System / Jie Lu. // Proceedings of the 2nd International Conference on Information Technology for Application (ICITA 2004). – 2004. – №2. – pp. 374–370. .
3. Бурая К.И., Виноградов П.Д., Грозин В.А. Автоматическая суммаризация веб-форумов как источников профессионально значимой информации / [Бурая К.И., Виноградов П.Д., Грозин В.А. и другие]/ Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 3. С. 482–496.
4. Барсегян А. А. Методы и модели анализа данных: OLAP и Data Mining / А. А. Барсегян, М. С. Куприянов. – СПб: БХВ-Петербург. – 336 с. – (БХВ-Петербург).\ ст.162.
5. Григорьев А. В. ОСНОВНЫЕ ПОДХОДЫ ПРИ СОЗДАНИИ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ / А. В. Григорьев, Е. А. Заплетин. // VI международная научно-техническая конференция студентов, аспирантов и молодых ученых (ИУСКМ-2015). – 2015. – №6. – С. 14–19.
6. K-Means Clustering in Python [Электронный ресурс]. – 2017. – Режим доступа до ресурсу: <https://mubaris.com/2017/10/01/kmeans-clustering-in-python>. (дата звернення: 02.11.2018)
7. Frey B. Clustering by Passing Messages Between Data Points / B. Frey, D. Dueck. // American Association for the Advancement of Science. – 2007. – №315. – pp. 972–976.
8. Sneha Y. A Study on Clustering Techniques in Recommender Systems / Y. Sneha, D. Mahadevan. // International Conference on Computational Techniques and Artificial Intelligence (ICCTAI-2011). – 2011. – pp. 97-100.

9. Kyoung-jae Kim, Hyunchul Ahn A recommender system using GA K-means clustering in an online shopping market / K. Kim, A. Hyunchul // Expert Systems with Applications. – № 34, issue 2. – February 2008. – pp. 1200-1209.
10. Pham M.C., Yiwei Cao, Ralf Klamma, Matthias Jarke A Clustering Approach for Collaborative Filtering Recommendation Using Social Network Analysis /M.C. Pham, Y. Cao, R. Klamma, M. Jarke// Journal of Universal Computer Science, vol. 17, №. 4. – 2011. – pp. 583-604.
11. Kuldeep S. Clustering of people in social network based on textual similarity / S. Kuldeep, K. S. Harish, B. Bhaskar. // Perspectives in Science. – №8. – 2016.– pp. 570—573.
12. Gurusamy V. Mining the Attitude of Social Network Users using K-means Clusterin / V. Gurusamy, S. Kannan, J. P. Regan. // International Journal of Advanced Research in Computer Science and Software Engineering. – May 2017. – №7. – pp. 226–230.
13. Miller Z. Twitter spammer detection using data stream clustering / [Z. Miller, B. Dickinsona, W. Deitrick and others.]. // Information Sciences. – 1 March 2014. – №260. – pp. 64–73.
14. Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-Based Social Networks / A.Noulas, S. Scellato, C. Mascolo, M. Pontil. // ICWSM. – 2011. – №11. – pp. 32–35.
15. Van Dam J., Den Velden Online profiling and clustering of Facebook users / J. van Dam, M. de Velden. // Decision Support Systems. – February 2015. – №70. – pp. 60–72.
16. Chen W., Fong S. Social network collaborative filtering framework and online trust factors: A case study on Facebook / W. Chen, S. Fong. // 2010 Fifth International Conference on Digital Information Management (ICDIM). – 5-8 July 2010. – №5. – pp. 266–273.
17. 2.3Clustering [Электронный ресурс] // Режим доступа: <https://scikit-learn.org/stable/modules/clustering.html#clustering>. (дата звернення: 05.11.2018)

18. Открытый курс машинного обучения. Тема 7. Обучение без учителя: PCA и кластеризация [Электронный ресурс] // Режим доступа: <https://habr.com/company/ods/blog/325654/>. (дата звернення: 05.11.2018)
19. Python official Documentation [Электронный ресурс] // Режим доступа: <https://www.python.org/doc/>. (дата звернення: 20.11.2018)
20. Django: The Web framework for perfectionists with deadlines [Электронный ресурс] // Режим доступа: <https://www.djangoproject.com/>. (дата звернення: 20.11.2018)
21. Современный учебник Javascript [Электронный ресурс] // Режим доступа: <https://learn.javascript.ru/>. (дата звернення: 23.11.2018)
22. jQuery [Электронный ресурс] // Режим доступа: <https://jquery.com>. (дата звернення: 23.11.2018)
23. HTML5 [Электронный ресурс] // Режим доступа: https://www.w3schools.com/html/html5_intro.asp. (дата звернення: 23.11.2018)
24. Bootstrap [Электронный ресурс] // Режим доступа: <https://getbootstrap.com/>. (дата звернення: 23.11.2018)
25. scikit-learn - Machine Learning in Python [Электронный ресурс] // Режим доступа: <https://scikit-learn.org/stable/>. (дата звернення: 20.11.2018)
26. Elasticsearch - RESTful search and analytics engine capable of solving a growing number of use cases [Электронный ресурс] // Режим доступа: <https://www.elastic.co/products/elasticsearch/>. (дата звернення: 20.11.2018)
27. PostgreSQL: The World's Most Advanced Open Source Relational Database [Электронный ресурс] // Режим доступа: <https://www.postgresql.org/>. (дата звернення: 23.11.2018)
28. MongoDB Atlas [Электронный ресурс] // Режим доступа: <https://www.mongodb.com/>. (дата звернення: 23.11.2018)
29. Касянчик Д. О. Використання методу k-середніх для об'єднання користувачів за даними з соціальних мереж / Д.О. Касянчик / Тринадцята міжнародна науково-практична конференція (МОДС. – 2018). – м. Чернігів.: ЧНТУ, 25 - 29 червня 2018 р. – С. 155-158

30. Касянчик Д.О. Інформаційна технологія кластеризації користувачів для підготовки цільових пропозицій / Д.О. Касянчик / Всеукраїнська науково-практична конференція молодих вчених та студентів «Інформаційні системи та технології управління» (ІСТУ-2018) – м. Київ.: НТУУ «КПІ ім. Ігоря Сікорського», 29-30 грудня 2018 р. – С. 16-19.

ДОДАТОК А

Графічний матеріал

ER діаграма

Математична модель методу підготовки цільових пропозицій користувачам з використанням нечіткої кластеризації

Схема алгоритму підготовки цільових пропозицій користувачам з використанням нечіткої кластеризації

Схема алгоритму роботи модифікованого методу кластеризації на основі методу поширення схожості

Результати експерименту

Структурна схема класів

Структурна схема розгортання